Business Applications of Machine Learning

Professor Kartik Hosanagar

Wharton
UNIVERSITY *of* PENNSYLVANIA

Agenda

- **Personalization on the Web**
  - Intro to recommender systems (content-based and collaborative)
  - Specific challenges of different systems
  - Addressing the challenges of personalization
- **Financial Applications**
  - Credit card fraud and machine learning in fraud detection
  - ML for identity verification, underwriting, churn predictions & more
- **Autonomous Vehicles**
  - Autonomous vehicles and the technologies they involve
  - Adoption challenges and potential

Personalization: Recommender Systems

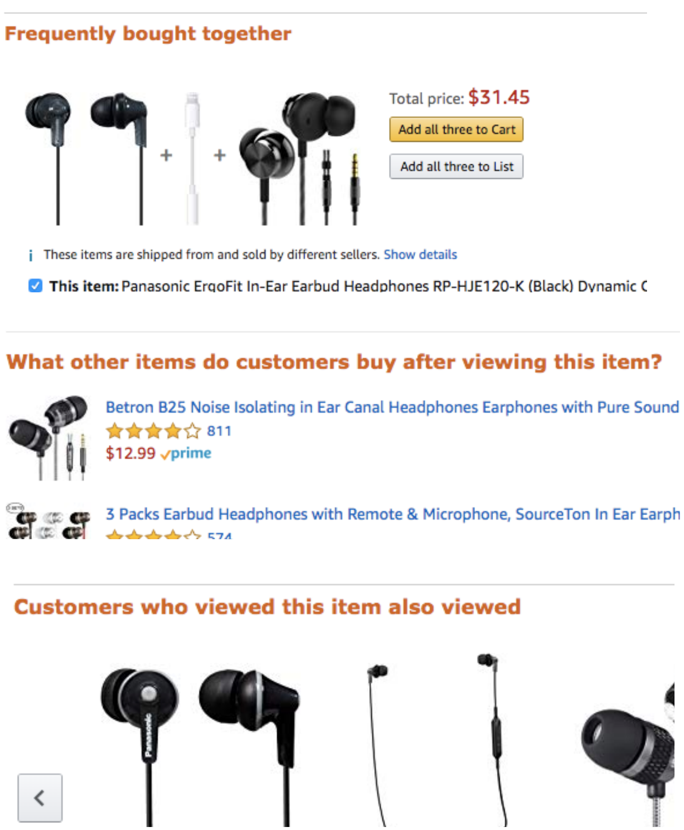
- Introduction to recommender systems
- Content-based recommenders
- Collaborative filters

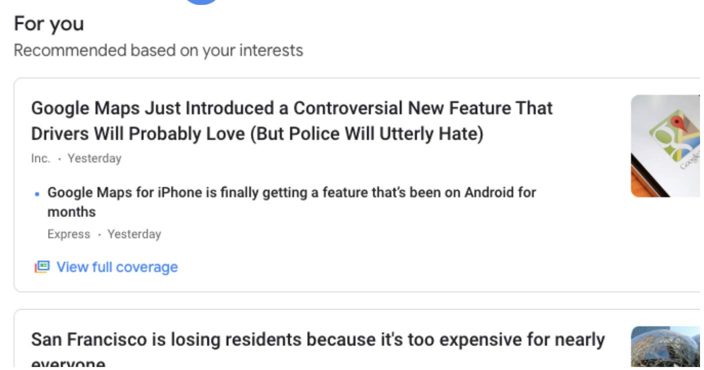
What are Recommender Systems

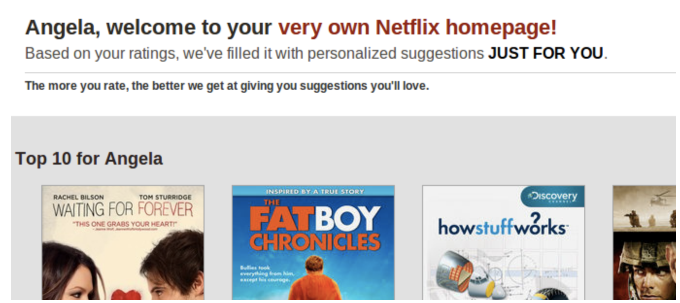
- Recommender systems use data on purchases, product ratings, and user profiles to predict which products are best suited to a particular customer.
  - Most common design is a ***collaborative filter***
    - “Customers who bought this item also bought…”
    - “People like you bought…”
- Recommender systems are valuable for both customers and firms.

Value to Customers	Value to Firms
• Learn about new products • Sort through a myriad of choices	• Convert browsers to buyers • Cross-sell • Increase loyalty

Recommender Systems Examples

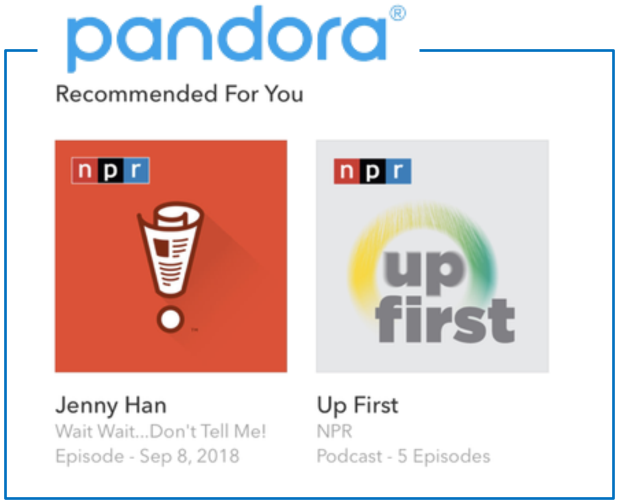




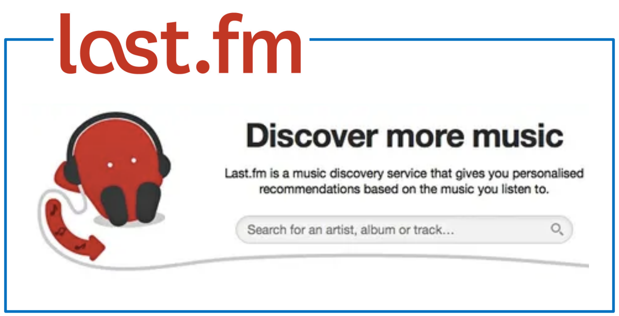


Two Main Recommender Designs

- **Content-Based Recommenders**
  - Find other products with similar attributes (Pandora)

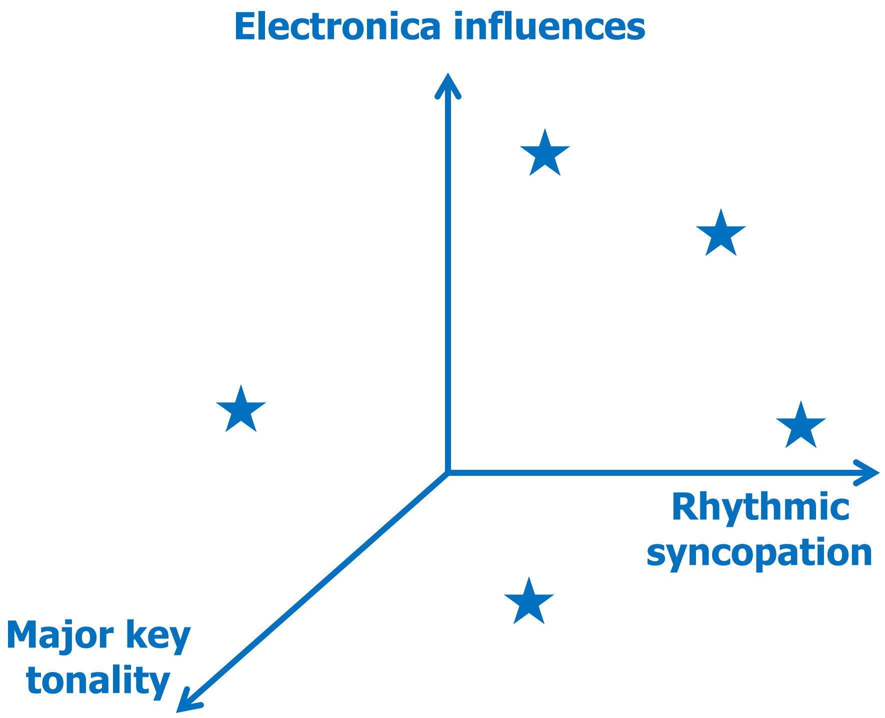


- **Collaborative Filtering (CF)**
  - Use information on what others buy/like (Last.fm, Amazon, Netflix)
  - "People who bought X also bought Y"



Content-Based Recommender: Pandora

- Once you find a song you like, Pandora finds other songs with similar musical qualities. Pandora emerged from the Music Genome Project, so it understands music & can provide explanations using accurate vocabulary.



Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Content-Based Recommender: Pandora

Pandora for Popular Genres

I ask for recommendations based on “Thunder” by Imagine Dragons.

↓

Pandora recommends tracks with similar musical qualities & explains why each is acoustically similar to “Thunder.” For example, “Ride” by Twenty One Pilots features “a dub production, a reggae feel, acoustic rhythm piano, use of a string ensemble and major key tonality.”

↓

I give one song I disliked a thumbs down.

↓

Pandora learns from this feedback & turns towards songs that are less similar to this song

Content/quotes from “A Human’s Guide to Machine Intelligence” by Kartik Hosanagar

Collaborative Filtering (CF)

- Not content-based, but rather based on what others with similar preferences are consuming
  - Amazon's "people who bought this also bought…".
  - Netflix's original design was also based on CF. It grouped users into "personas" & made suggestions based on users of similar "personas."

Last.fm for music recommendations

I ask for recommendations based on "Thunder" by Imagine Dragons.

↓

Last.fm suggests tracks based on what others who like "Thunder" also liked. Last.fm can't explain further b/c it lacks musical knowledge.

Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Collaborative Filtering (CF): Last.fm

- Two types of collaborative filters:
  - Item-to-Item CF: recommends items bought by others who bought the item you are interested in. ("People who bought X also bought Y")
  - User-similarity based CF: recommends items bought by others who are similar to you based on data the company has about your preferences ("People like you like X").
- Pros of collaborative filters:
  - Easy and cheap to build (don't need detailed metadata about products)
  - Effective in practice

Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Collaborative Filter Challenges

- Key design trade-offs
  - **Sparse data:** How do you decide how similar two users are when the intersection of products they have both used is limited? They may have each only reviewed a handful of products out of thousands, and may not have reviewed any product in common.

	1	2	3	4	5	6	7	8	9
U1	-	-	-	-	3	-	-	-	-
U2	-	-	-	-	-	-	-	-	4

← Here, U1 and U2 each rated a different product, making it difficult to tell how similar they are.

  - **Cold-start:** How to make recommendations for new users for whom you have no data? How to handle new items with no reviews or other users?

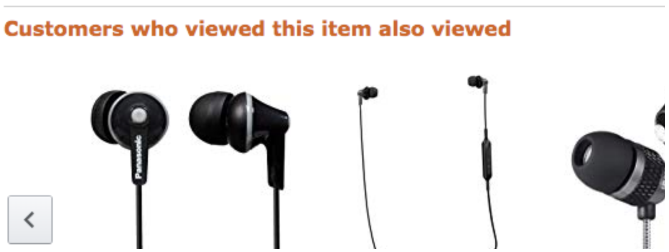
Content/quotes from “A Human’s Guide to Machine Intelligence” by Kartik Hosanagar

Personalization: Impact on Markets

- Specific challenges of collaborative filters
- The “Long Tail” and sales diversity

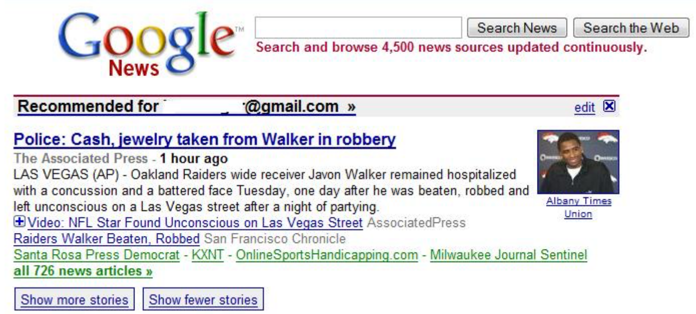
Recommenders and Consumer Choice

amazon



28-35% of retail sales at Amazon originate from algorithmic recommendations

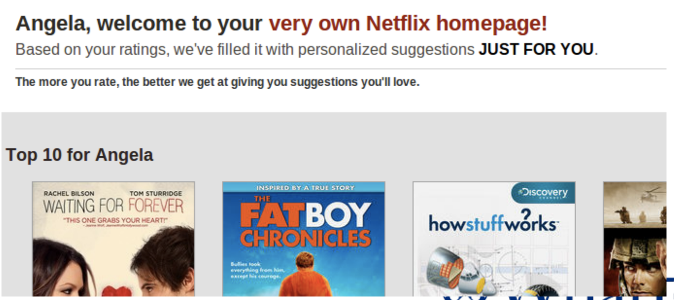
Google



38% increase in usage at Google News

NETFLIX

80% of video hours streamed on Netflix originate from algorithmic recommendations



Wharton UNIVERSITY of PENNSYLVANIA

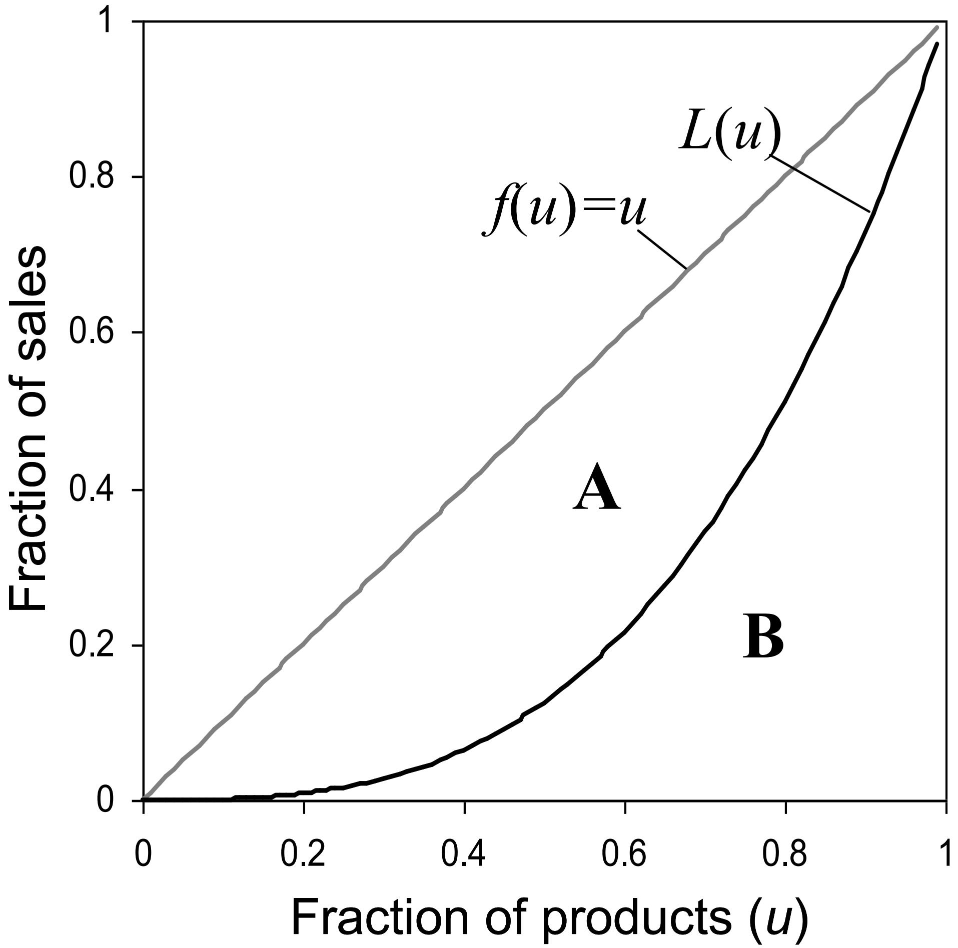
Collaborative Filtering and Sales Diversity

- The "Long Tail" concept: popularized by the book *The Long Tail*
  - "Suggested that the main effect of automated recommendations would be to help people move from the world of hits to the world of niches- obscure products that are closer to our individual preferences but never get our attention in mainstream markets."
- But given that CF algorithms recommend items based on what others are consuming, do they really increase sales diversity?

Source: Longtail.com

Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Measuring Diversity: Gini Coefficient



G = A/(A+B)

Closer to 0 = Equality

Closer to 1 = Inequality

Wharton
University of Pennsylvania

Problem Statement

- G_0 = Gini without recommendations
- G_i = when recommender r_i is employed, all else equal

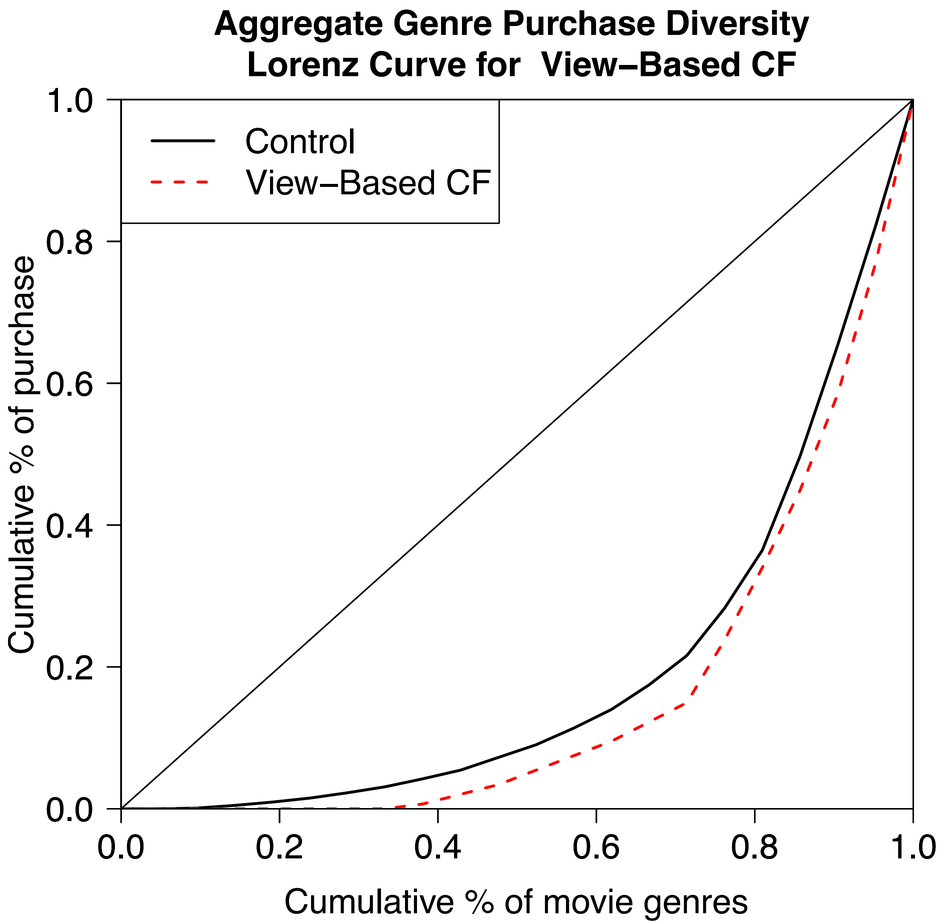
Bias of r_i	**Diversity**	$G_i < G_0$
	Concentration	$G_i > G_0$
	Neutral	$G_i = G_0$

Wharton
UNIVERSITY of PENNSYLVANIA

Empirical Investigation

- Field experiment on a top retailer website
  - August 8, 2013 and August 22, 2013
  - A/B test with visitors given unique IDs and all behavior tracked over time
  - 1,138,238 users randomly assigned into 3 groups
    - **1. Control (77%)**
    - **2. View-Based Collaborative Filtering (11.5%)**
      - "People who viewed this item also viewed"
    - **3. Purchase- Based Collaborative Filtering (11.5%)**
      - "People who purchased this item also purchased"

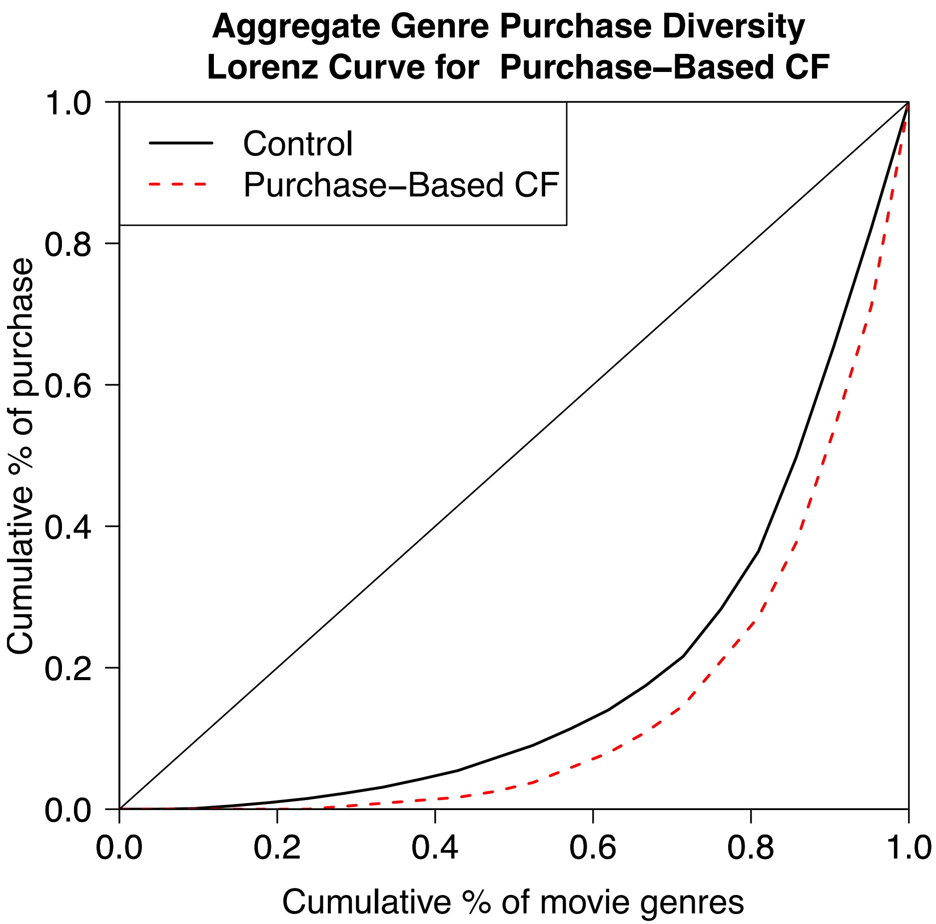
Sales Diversity of Movies*



	Control	View-based CF
Gini	0.60	0.68
Change |p-value		0.07 | (0.004)

* Results are directionally similar for all other product categories on the website

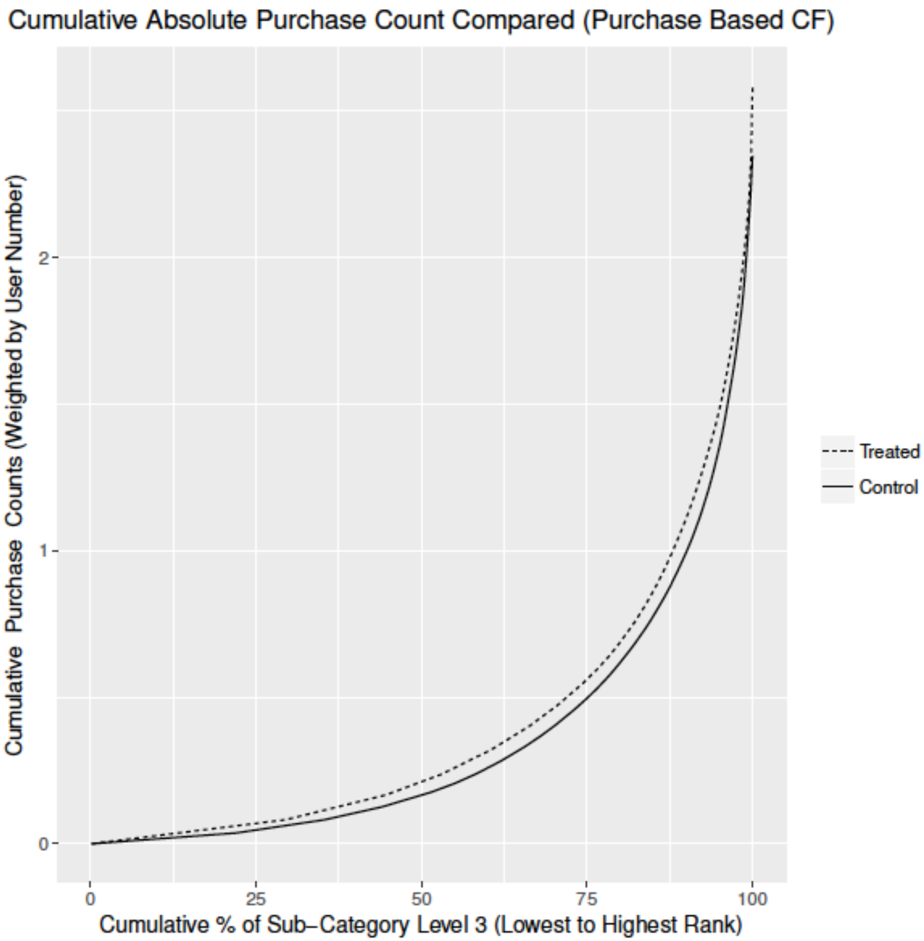
Sales Diversity of Movies*

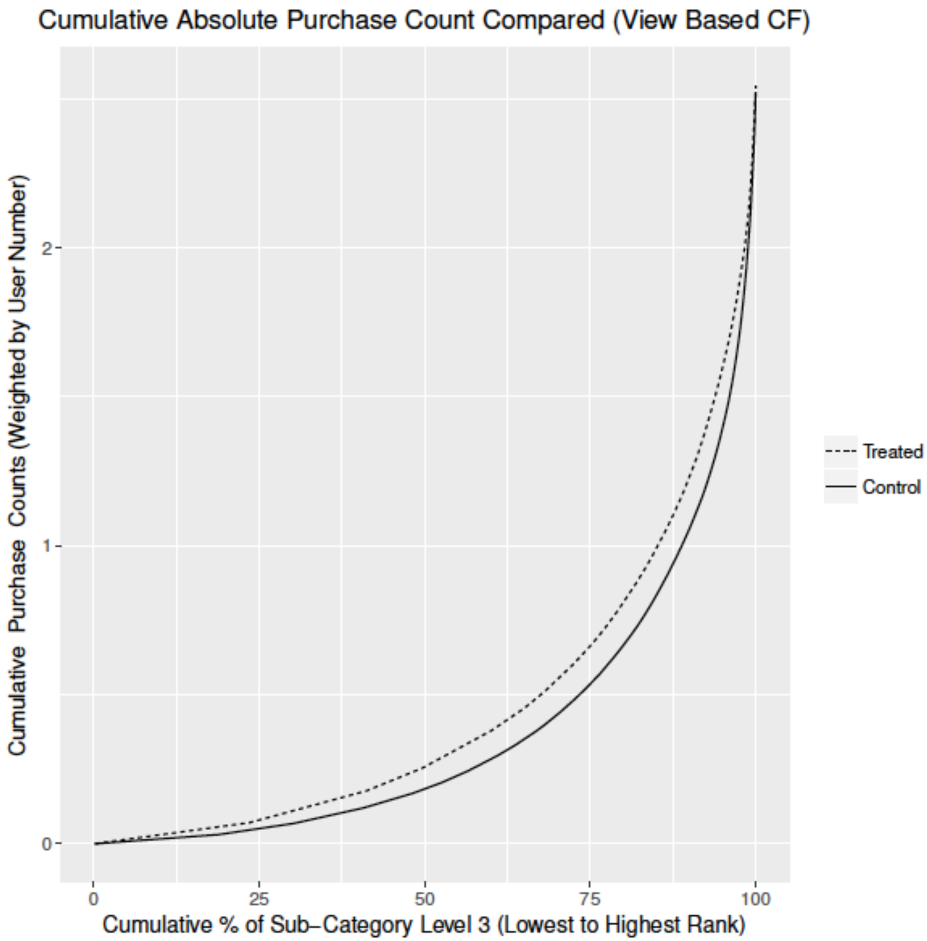


	Control	View-based CF
Gini	0.60	0.70
Change |p-value		0.10 | (< 0.0041)

* Results are directionally similar for all other product categories on the website

Absolute Sales of Niche Items





Wharton
UNIVERSITY of PENNSYLVANIA

Absolute Sales of Niche Items

Even though the market share of niche items decreases after the firm rolled out traditional collaborative filtering algorithms, the absolute sales of niche items in fact increases. That is, all products – niche and popular – sell more under collaborative filters. But popular products gain much more than niche products, resulting in their market share further increasing.

Wharton
UNIVERSITY of PENNSYLVANIA

Personalization: Addressing the Challenges

- Tradeoffs of different recommender designs
- Hybrid recommenders
- Moving beyond recommender systems
- Risks of personalization

Tradeoffs of Different Recommender Designs

Collaborative Filtering

PROS

- Helps users discover new items
- Knowing that others "like you" like an item can itself be convincing
- No need for detailed metadata

CONS

- Has a popularity bias
- Can't explain recommendations
- Has difficulty with less popular items that have less data
- Cold-start issue (new items)

Content-Based Recommenders

PROS

- Provides relevant suggestions
- Can explain recommendations
- Works relatively well even for less popular/newer items with less data as well as for new users.
- Doesn't have popularity bias

CONS

- Difficult to build because you need detailed metadata

Hybrid designs combine the simplicity of CFs with the "fairness" of content-based designs.

Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Hybrid Recommender: Spotify

- Spotify Discover uses a hybrid design and addresses the difficulty of needing detailed metadata by applying ML to automate this process.
  - To gain the knowledge of a content-based recommender, Spotify:
    - "Crawls the web to examine blog posts & online discussions to figure out the kind of descriptive language that listeners use to discuss songs/artists. It then uses these terms for attributes of songs."
  - But there is less data/discussion about newer and niche songs online
    - Spotify "uses ML to analyze the audio signal of a song and extract characteristics like tempo, loudness, key & tonality."
- Spotify Discover combines the best of both types of recommenders, uses ML to automate the work of building the system's musical knowledge, avoids a popularity bias, and is very well liked by users.

Content/quotes from "A Human's Guide to Machine Intelligence" by Kartik Hosanagar

Personalization: More than Recommendations

- Personalization is not limited to product recommendations. It is about holistically adjusting communications with customers based on customer characteristics (e.g. websites or emails tailored to individual users).
  - If you know a customer is located in Seattle, you might send them an ad for rain gear, while a customer in Boston may get one for snow gear.

Wharton
UNIVERSITY of PENNSYLVANIA

Risks of Personalization

- **Misapplication**
  - Making sweeping generalizations
- **Data privacy concerns**
  - It isn't always effective to remind customers how much info you have on them
- **Crossing the "creepy" line**
  - I may not want a website to remember that I've bought weight loss products or smoking-cessation products
- **Regulatory Compliance**
  - Respect laws regarding data privacy in multiple jurisdictions