On the Economics of Recommender Systems

Emilio Calvano Center for Studies in Econ and Finance U. Napoli Federico II

June, 2015

Recommender Systems defined

- Recommender Systems are software tools and techniques providing suggestions for items to be of use to a consumer.
- Support users in various decision-making processes, such as what items to buy, what music to listen, or what news to read. Recommender systems have proven to be valuable means for online users
- cope with information overload / abundance of choice
- hugely powerful and popular tools in electronic commerce.
- Amazon, Netflix, OK Cupid, Pandora...

On Netflix...

Ranking
Rows
Netflix
Recently Watched
Top 10 for Justin
Popular on Facebook
Goofy TV Shows
Visually-striking Exciting Foreign Movies
Sci-Fi & Fantasy

Estimate a utility function that automatically predicts how a user will like an item

Xavier Amatriain
(Engineer - director of algorithm engineering at Netflix)

Bssed on

- Past behavior
- Relation to other users
- Item similarity
- Context
- ...

Do they actually Shape consumption choices?

Relevance

What do people (I talk to) say:

- 'Yes - totally obvious'
- 'No - I don't pay attention / I know what I want'

What do companies say:

- Netflix: 2/3 of the movies watched are recommended (Xavier Amatriain - Engineer Director for the Algorithms Engineering team at Netflix)
- Google news: recommendations generate 38 % more clickthough
- Amazon says 35 % of product sales result from recommendations. - (Matt Marshall, VentureBeat)

What do we know? (empirical / experimental evidence)?

- OK Cupid filtering on 'interactions'
- Facebook filtering on news consumption (Athey and Mobius (2015))

OK Cupid experiment (1)

Does the 'match' algorithm work?

- Idea: Switch off the recommendations and see the outcomes.
  - e.g. you tell couples they are 'bad' matches regardless of their 'predicted' compatibility.

Odds of a single message turning into a conversation

		number DISPLAYED to them		
		30% match	**60% match**	**90% match**
ACTUAL compatibility of users	**30% match**	10%	16%	17%
	60% match	13%	13%	16%
	90% match	16%	17%	20%

OK Cupid experiment (2)

Does it shape consumption?

- Idea: Randomize recommendations and see the outcomes.
    - e.g.
    - Idea: take predicted ‘bad’ matches and tell them they are ‘good’ matches. compatibility.

Odds of a single message turning into a conversation

		number DISPLAYED to them		
		30% match	**60% match**	**90% match**
ACTUAL compatibility of users	**30% match**	10%	16%	17%
	60% match	13%	13%	16%
	90% match	16%	17%	20%

Why shall we care? (from a public policy perspective)

Why shall we care?

1. Fresh means to exhert market power?
   - What is an abuse of a dominant position (art. 102)?
   - What is "unfair" or works to the "prejudice" of consumers here?
2. New grounds for anti-competitive practices?
   - Big Data as a barrier to entry.
3. Privacy and consumer protection issues
4. Promote (ideological) Diversity / Filter bubbles and echo chambers
   - Concern: "Personalized" rec. $\rightarrow$ **'algorithmical segregation'**
     - Always Listen to same music, watch similar movies, exposed to same ideology...
     - RecSys 'reinforce' existing taste / don't expose users to new ones
     - In fall 2014 France's Council of State recommended government oversight over the algorithm that Netflix uses to present series and movies, to make sure French and European content is well positioned.
   - Opportunity: 'Serendipitous algos' (i.e. algos delivering pleasant surprises) are rewarded by the market and therefore developed.

Recommendation bias: (a few) insights from theory

To fix ideas

- A set of objects (say: movies)
- A Consumer (hereafter C) who can't tell the objects apart
- A Recommender (RS) who
  - has a technology to predict taste.
  - can recommend / not based on prediction.

- Naive intuition suggests that NETFLIX always recommends the 'best' (i.e. CS maximizing) movie.
- In what follows I speculate about potential potential wedges between RS and C incentives

1. Financial Incentives
2. 'Surplus extraction' incentives
3. Reputational incentives (Calvano and Jullien (2015))

1 - Financial incentives

Well understood: RS may have preferences over what consumers choose:

- kickbacks, commissions, heterogeneous margins, 'own' content House of cards, Amazon branded product, Google shopping

Not so well understood: why are these contractual incentives there?

- Right Conceptual framework: vertical chain.
- RS are often bottleneck suppliers of attention.
- RS are often akin to big downstream retailers.
- Usual Chicago critique calls for ad-hoc foundation of the recommendation bias.

On Netflix and its business metrics

Hugely popular DVD rental company (now mostly streaming).

- 50M subscribers.
- 7B hours/quarter.
- 90 minutes a day (Avg)
- 150 choices (clicks) a day.

Their metrics

- Retention of existing customers
  (fraction of subscribers who renew subscription)
- Creation of new ones

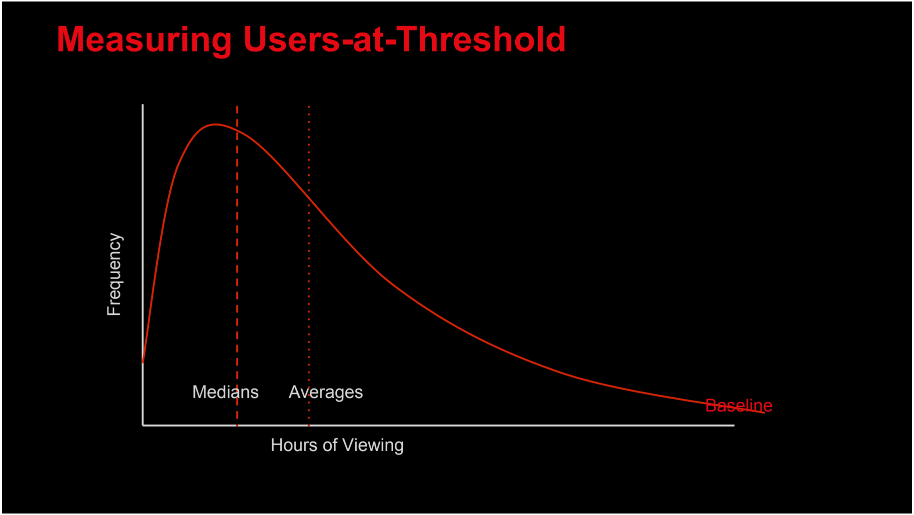
Their biggest challenge: customer retention.

- 0.1% increase in retention $\approx$ \$50M / year

- Outsource research (Netflix Contest)
- Large scale experimentation with different algos
  - A/B testing with more than 500k users per cell.
- Use customer retention (or other obvious predictors such as #hours watched) to asses the alogs.

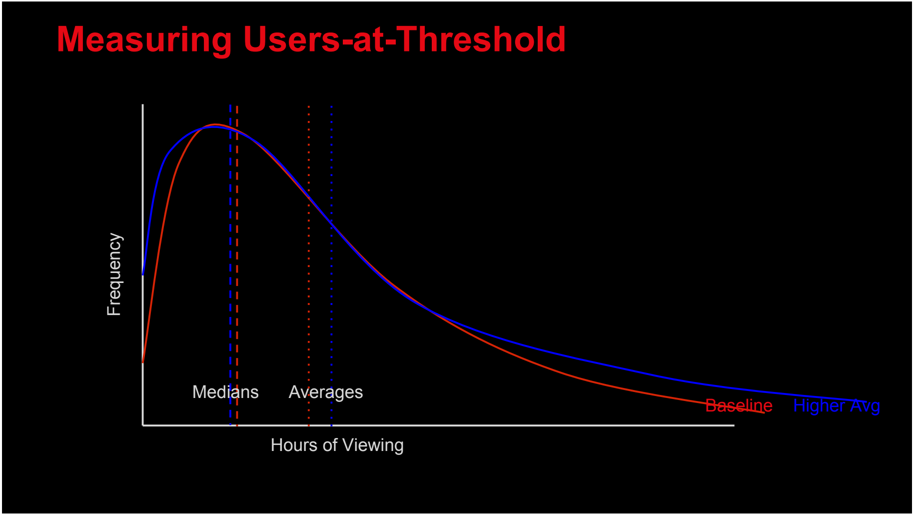
What is wrong with that?

One (revealing) experiment...



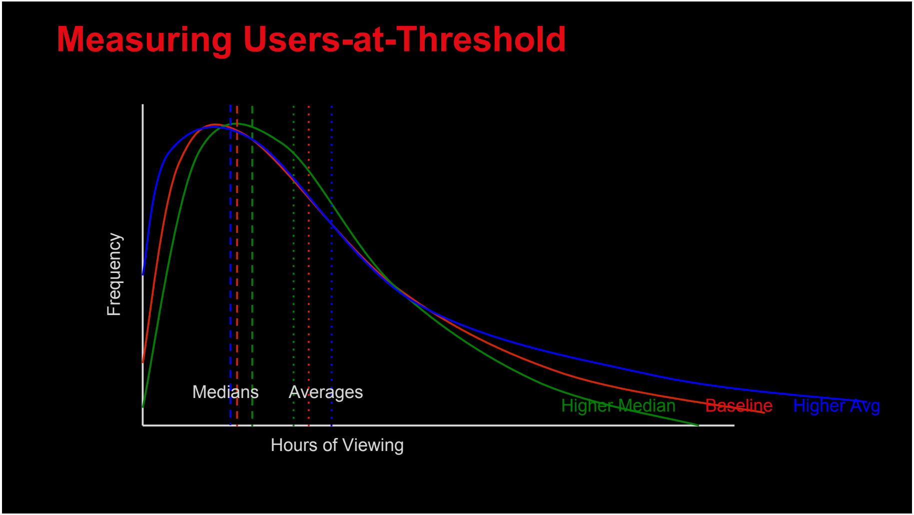
Source: N. Hunt (2014)

One (revealing) experiment...



Source: N. Hunt (2014)

One (revealing) experiment...



Source: N. Hunt (2014)

PUNCHLINE...

Netflix Caters to marginal consumer not average;
a.k.a the Spence distortion

NEIL HUNT - CHIEF PRODUCT OFFICER AT NETFLIX - OCTOBER 2014

All the work we do to make better recommendation [. . .] is basically testing to see whether this one key person [on the fence between renewing or not the subscription] falls on this side of the fence or the opposite side.

- Algo is biased towards marginal viewer.
- Vivid illustration of the 'Spence' distortion.
- 'Awareness' is not a necessary ingredient: A/B testing with the right metric does the trick.

A (MOTIVATING) QUESTION:

Consumers cancel their Netflix subscription more often after:

1. a stretch of bad movies
2. a stretch or good movies
3. It doesn't matter

A (MOTIVATING) QUESTION:

Consumers cancel their Netflix subscription more often after:

1. a stretch of bad movies
2. a stretch or good movies
3. It doesn't matter

- Recommendations are experience goods
- Individuals assess (make inference) the value from staying hooked up (i.e. subscribed) to Netflix.
- Bad movies signal bad news about the 'quality' of the service (that is intentionally vague)

A toy model of Netflix (Based on Calvano and Jullien (2015))

- One recommender ((N)etflix) and one consumer (C)
- Two periods (say: months).
- One new object (movie) every month.

Every period:

- Netflix chooses to recommend or not the movie.
- C follows advice and then chooses to renew subscription or not.

Netflix Basic goal:

- 'Persuade' C to renew subscription at the end of month 1.

To make the problem interesting...

- Assume C renews

- Movie is either , , , , 

Public information

- (Average) star rating of the other subscribers (prior)
- The opportunity cost of C's time is 

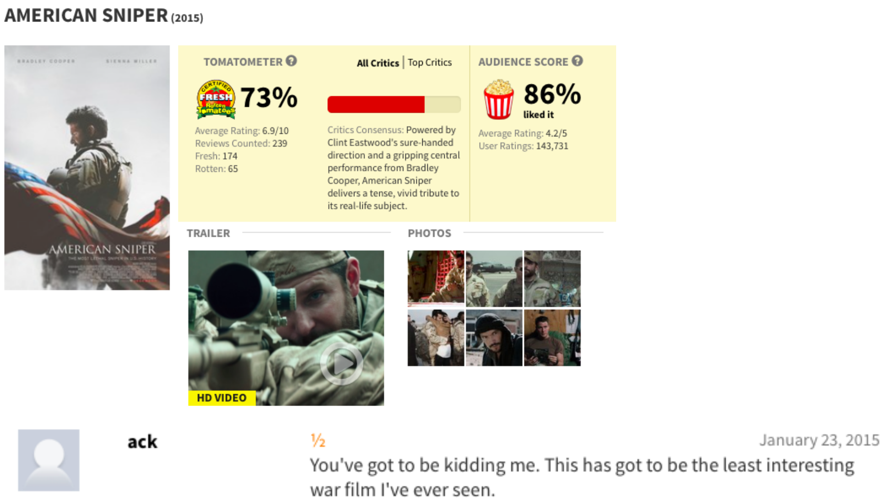
Private information (key)

- Netflix can be one of two types: Clueless  / Oracle 

- Oracle observes (almost) actual taste.
- Clueless observes (almost) nothing.
- Common prior.

Consumer problem: Figure out Netlix's 'type' after watching movie.

Average rating and taste heterogeneity



Average rating and taste heterogeneity (2)

DUMB AND DUMBER TO (2014)

TOMATOMETER

29%

Average Rating: 4.3/10
Reviews Counted: 134
Fresh: 39
Rotten: 95

All Critics | Top Critics

Critics Consensus: Dumb and Dumber To does have its moments, but not enough of them -- and the Farrelly brothers' brand of humor is nowhere near as refreshingly transgressive as it once seemed.

AUDIENCE SCORE

38% liked it

Average Rating: 2.9/5
User Ratings: 64,259

Edward T ★ May 31, 20

You have to be dumb to sit through this drivel.

Lindsey C ★★★★½ May 31, 2015

Loved it!!! Had me belly laughing from the start.

A negative result

Suppose that C 'naively' believes Netflix is 'honest.'

- N Recommends movie only if its 'best guess' > ★★★

Is honesty the best policy? **No**

INTUITION VIA EXTREME ILLUSTRATION

To make problem meaningful: assume C renews **only if** RS is an Oracle.
Suppose movie 'very good' (ex-ante): Average rating: ★★★★

Consumer reasoning

- If N recommends and movie 'sucks' ($<$ ★★★) then N must be clueless $\rightarrow$ stop subscribing.
- If N recommends and movie 'good' ($>$ ★★★) then N could be either clueless or oracle: stop subscribing.
- if N does not recommend then must be Oracle $\rightarrow$ keep on subscribing.

Netflix reasoning: regadless of type: do not recommend.

3 LESSONS FROM CJ (2015) (AND COUNTING...)

1. The need to $\uparrow$ retention $\rightarrow$ strategic rec. bias
   - distort rec. inefficiently *away* from the prior
   - Over-recommend bad movies
   - Under-recommend good movies
2. Strategic incentive is self-defeating
   - Only pooling eqm can coexist
   - Not much information gets transmitted.
   - Under-recommend good movies
3. There is no obvious market reward for using information in the 'right' way:
   - RS tends to 'conform' to C's expectations (minimize disappointment)
   - competition may increase sensitivity to reputation $\rightarrow$ exacerbate bias.