

Suggesting Watches:

A Search-Optimality Experiment on the Effect of Video Recommendations

Walter Stover

Abstract:

Consumers increasingly interact with recommender systems when making consumption choices, including media consumption on digital platforms such as YouTube. Experimentally testing the effect of recommender systems on consumer choice faces the difficulty of ranking media consumption alternatives which have little to no technical basis for comparison. We address this by treating media consumption on digital platforms as a search problem, enabling us to generate a list of alternatives from a probability distribution and allowing calculation of optimal search stopping points. We use this method of generating alternatives to simulate the choices consumers face on digital platforms such as YouTube. We then conduct an experiment testing the effects of recommender systems on search optimality by having participants engage in search through the generated alternatives with and without the presence of a recommender.

Introduction:

A significant portion of all online experience appears to us ordered by recommendation agents, from online commerce (Amazon), and more recently, entertainment platforms such as YouTube. These websites are all characterized by a very large number of options for consumers to choose from – in the case of YouTube, approximately 30,000 hours of video are uploaded every day – an amount requiring an individual to spend [82 years](#) to watch all the content uploaded in just a single hour. This is winnowed to just a handful of videos that appear on your homepage for you to glance at.

This has led to myriad questions over how the recommendation algorithm ends up deciding what videos to select for the homepage, as well as raising concerns over information overload, algorithmic overdependence, harmful echo chambers, and even radicalization. However, investigating any of these questions require us to be able to say something about the causal effect of these recommendations on the choices of consumers and the optimality of those choices. This is difficult; for one, rankings are endogenous: digital platforms want to rank the most relevant choices highest, but consumers will pay more attention to the highly ranked choices (Ursu 2018). Second, digital platforms involve choices between alternatives that have no basis for objective technical comparison like you could have when shopping for tangible goods; aside from the problem of consumer subjective preferences, you also can't say that one YouTube video is objectively better than another from a price-point analysis like you could about phone chargers on Amazon (Banker and Khetani 2019).

In this paper, we propose investigating the effects of video recommendation algorithms on consumer welfare by conceptualizing video selection as an optimal search problem. Consumers only have limited information about each video that they see on their homepage, and then receive a probabilistically distributed payoff from watching that video. The search problem from the consumer's side is how many videos the user should watch before they terminate their search process. The problem from the

recommendation algorithm side is how best to arrange videos for consumer search based on the payoffs incurred by the consumer in the past.

Literature Review:

This paper is motivated by the experimental literature on the effects of recommenders on consumer behavior. Seminal papers by Haubl & Trifts (2000), West et al. (1999) showed benefits in the form of reducing consumer search efforts, increasing quality of the sets of options for consideration, and improving quality of purchase decisions. Later papers also have focused not just on individual decision quality, but also on the effects of recommenders on aggregate consumption choices (Fleder and Hosanagar 2009). These papers and the rest of the literature focus primarily on online shopping environments – for instance, further papers by Pathak et al. (2010) on the effect of recommenders on sales; Kim, Albuquerque and Bronnenberg (2010) looking at Amazon data; and Banker and Khetani (2019) using Amazon products to test for algorithmic overdependence.

However, little work has been done in extending these experiments to treat online entertainment providers such as YouTube. Many of these platforms, including YouTube, Spotify, Netflix, etc., are examples of digital platforms, with “...the core mission of enabling and generating value from interactions between users” (Belleflame and Peitz 2020). Related literature on digital platforms from a cultural economic perspective is either non-experimental, such as Peukert (2019), Hosanagar et al. (2014), and Kretschmer and Peukert (2018) which specifically looks into YouTube; or it’s not directly experimental as Tucker and Zhang (2011) which analyze data from a prior field experiment run by a website. This literature is also often concerned with recommenders together with ratings and reviews, known as 3R systems, rather than focusing specifically on recommender systems.

There’s also a related body of experimental literature on digital platforms coming from the MovieLens database, as in Chen et al. (2010). However, this doesn’t look at recommenders but at the effect of information provision on user ratings. This is a critical input into recommender systems, but does not itself specifically examine the effect of the recommenders on the quality of consumer decisions thereafter. There are MovieLens field experiments conducted such as Ekstrand et al. (2015), but these are non-theoretical papers that do not attempt to quantify the causal effect of recommender system choice or recommender attributes on users.

Our paper thus fills several niches in the bodies of literature thus provided. First, we are extending the experimental treatments of recommender systems to digital platforms beyond the original marketing experiment corpus of shopping platforms. Second, we are extending the cultural economics literature on digital platforms by introducing experimental methodologies. Third, we are extending the experimental literature on digital platforms by focusing specifically on the effects of recommender systems.

We propose to address these gaps by using Weitzman (1979) as a method of generating search alternatives, enabling precise ranking and calculation of search optimality. Weitzman (1979) has been used before in related search literature, such as Kim, Albuquerque and Bronnenberg (2010), Honka and Chintagunta (2017), and Ursa (2018). However, these are not explicitly experimental papers; instead, they “take Weitzman to the data” and use this approach to quantify the causal effects of rankings from existing data. Ursu (2018) does use Weitzman to estimate causal effects of ranking in the context of an experiment, but a field experiment previously run by Expedia. Kim, Albuquerque and Bronnenberg

(2010) use numerical experiments that simulate agents and their rankings as supporting evidence, but do not involve real agents in the experiment. Finally, Honka and Chintagunta (2017) has no relationship to experimental procedures at all. We believe that using Weitzman (1979) to generate simulated search alternatives is a novel approach that will allow experimental testing of the effects of recommender systems in digital platform contexts where it's ordinarily difficult or impossible to rank search alternatives.

Theory:

We want to examine the effects of recommenders in the context of video recommendations on digital platforms such as YouTube – but estimating the causal effects of recommendations is difficult, because we cannot determine what videos are technically dominant to use as a benchmark as in Banker & Khetani (2019). To overcome this, we conceptualize the problem of consumer selection of videos as a search process using the optimal search process theory of Weitzman (1979). Weitzman uses the concept of a *reservation price*, which he defines as an “invariant critical number analogous to an internal rate of return” that depends only on the features of the source of a reward. He then leverages the concept of a reservation price to formulate a *selection rule* and a *stopping rule*. The selection rule is to search next for the unsampled source with the highest reservation price, and the optimal stopping point is when maximum sampled reward is above the reservation price of every unsampled source.

The reservation price is defined as $z_i = \frac{p_i R_i - c_i}{p_i}$, which simply states that the reservation price of a “box” (a source of reward) is the expected net gain distributed according to some probabilistic function, divided by the probability of success.

This reservation price design solves a very hard problem with testing recommendation agents in terms of the quality of consumption, especially in the context of platforms centered on consumption of media, such as video or music recommendation platforms. As noted by Haubl & Trifts (2000) in the context of recommendation agents in online marketplaces, “Measuring the *quality* of purchase decisions and consideration sets is a very ambitious endeavor... [b]ecause an individual’s preferences are not subject to direct observation, it is impossible to accurately measure decision quality in uncontrolled real-world settings.”

Weitzman (1979) allows us to pursue an alternative path where, instead of conducting an experiment with existing videos that we would have to try and calculate the value of, we can simply create our own distribution of “videos”. These videos are just drawn from a distribution with realizations of some payoff and some cost. The search process for the highest-value video can then be abstracted by simulation using these videos with a randomized reservation price which is determined by the realized payoff and cost. Since we know the reservation price for all the videos, we can then determine the exact payoff at any point in the search process, as well as form selection and stopping rules for an optimal search process.

Experimental Design:

Commented [WS1]: Counter: is there data generated by YouTube that we could take Weitzman to?

Drawing on this discussion of using reservation prices to simulate a search process, we construct an experimental design where subjects will engage in a search process in a distribution of videos, each with their own reservation price. This experiment will consist of three treatments:

1. Treatment 1 is a baseline treatment with a Weitzman list of search alternatives and no recommendation.
2. Treatment 2 will have a list of search alternatives and then randomly recommend a video each round.
3. Treatment 3 will recommend high-value videos in each round.

In each treatment, subject participants start with a budget of \$20. Each video costs a randomized amount to watch, with a randomized reward unknown to the participant until they decide to watch it. After each video is watched, the participant may decide to keep searching and watch another video, or they may decide to end with their current earnings. If the subject ends the round, they immediately move onto the next round. After all rounds are over, if the subject earned a positive amount above budget, they are paid that total amount. For instance, if across all ten rounds, the subject spent \$15 on search, but gained \$30 in payoffs, then they would end up with a net \$35 payoff, as they would end with \$15 higher than their starting budget of \$20. Subjects can lose down to the show-up fee, so they will be paid down to \$10 as a floor; if subjects end with a lower budget than this, they will still be paid the show-up fee.

This should provide sufficient incentive for participants to engage in the search process. The search process will consist of selecting videos to watch, with expectations formed by summarized information regarding the relevant probabilities of payoffs and costs, and to then subsequently decide if they have made enough. Users must also watch at least two videos per round to receive the show-up fee, so they can't just skip all the rounds to the end.

The experiment will proceed by having participants log onto our platform in the computer lab, where, after having the experiment explained to them, they can start the treatment that they have been randomly assigned to. The platform will display a square grid layout of videos which will be simple boxes with nothing but the watch time label inside the box. It will also display their remaining budget, as well as a simplified summary of the probabilistic distribution of payoffs for the videos available on the grid.

Data Generation:

For the purposes of this class, I will use an Agent-Based Model (ABM) to generate my own data. I will use simple learners endowed with heterogeneous levels of trust in a recommender system. Agents will have summary information about the distribution of reservation prices each round, but will not know the exact price of any given video. Agents will engage in simple spatial search where they will randomly select a video and then try videos in the Von-Neumann neighborhood east, west, north, and south of that video. If the video is revealed to be more valuable than the one they had watched previously, the agent will "move" to this new area of the grid and continue searching. Each agent has heterogeneously endowed search preferences that determine how much search cost they are willing to undergo before they will stop the search process.

In treatments where a recommender system is present, agents will have a % chance of ignoring their own search process and selecting the recommended video instead, determined by their endowed level of trust in recommender systems.

This should be a simple enough ABM to be feasible to code for this class, and will generate enough data to then conduct econometric analysis, for instance, with the presence of recommender systems as dummy variables, and search optimality as the variable of interest as defined by how far off agents were from the optimal stopping point at the time that they terminate search.

Formula:

For the control, we don't have a linear regression, but simply measure the search quality directly.

For the first treatment, our formula will be the following:

$$SearchQuality = \beta_0 + \beta_1 Recommend + \epsilon$$

Where *Recommend* is a dummy variable indicating the presence of recommendations.

For the second treatment, our model is the following:

$$SearchQuality = \beta_0 + \beta_1 RecommendQuality + \epsilon$$

Where *RecommendQuality* refers to the quality of the recommendation – i.e., what percentile of the highest video values is the recommender system picking from in the distribution.

Bibliography

- Banker, Sachin, and Salil Khetani. "Algorithm Overdependence: How the Use of Algorithmic Recommendation Systems Can Increase Risks to Consumer Well-Being." *Journal of Public Policy & Marketing* 38, no. 4 (October 1, 2019): 500–515. <https://doi.org/10.1177/0743915619858057>.
- Belleflame, Paul and Martin Peitz, "Ratings, reviews and recommendations." In *Handbook of Cultural Economics*, edited by Ruth Towse and Trilce N. Hernandez, 466–471. Cheltenham, U.K., Northampton, MA, USA, Edward Elgar Publishing, 2020.
- Chen, Yan, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li. "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens." *American Economic Review* 100, no. 4 (September 2010): 1358–98. <https://doi.org/10.1257/aer.100.4.1358>.
- Ekstrand, Michael D., Daniel Kluver, F. Maxwell Harper, and Joseph A. Konstan. "Letting Users Choose Recommender Algorithms: An Experimental Study." In *Proceedings of the 9th ACM Conference on Recommender Systems*, 11–18. RecSys '15. New York, NY, USA: Association for Computing Machinery, 2015. <https://doi.org/10.1145/2792838.2800195>.
- Fleder, Daniel, and Kartik Hosanagar. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity." *Management Science* 55, no. 5 (May 2009): 697–712. <https://doi.org/10.1287/mnsc.1080.0974>.
- Häubl, Gerald, and Valerie Trifts. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids." *Marketing Science* 19, no. 1 (February 2000): 4–21. <https://doi.org/10.1287/mksc.19.1.4.15178>.
- Honka, Elisabeth, and Pradeep Chintagunta. "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry." *Marketing Science* 36, no. 1 (January 2017): 21–42. <https://doi.org/10.1287/mksc.2016.0995>.
- Hosanagar, Kartik, Daniel Fleder, Dokyun Lee, and Andreas Buja. "Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation." *Management Science* 60, no. 4 (April 2014): 805–23. <https://doi.org/10.1287/mnsc.2013.1808>.
- Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg. "Online Demand Under Limited Consumer Search." *Marketing Science* 29, no. 6 (November 2010): 1001–23. <https://doi.org/10.1287/mksc.1100.0574>.
- Kretschmer, Tobias, and Christian Peukert. "Video Killed the Radio Star? Online Music Videos and Recorded Music Sales." *Information Systems Research* 31, no. 3 (September 2020): 776–800. <https://doi.org/10.1287/isre.2019.0915>.

- Pathak, Bhavik, Robert Garfinkel, Ram D. Gopal, Rajkumar Venkatesan, and Fang Yin. "Empirical Analysis of the Impact of Recommender Systems on Sales." *Journal of Management Information Systems* 27, no. 2 (October 1, 2010): 159–88. <https://doi.org/10.2753/MIS0742-1222270205>.
- Peukert, Christian. "The next Wave of Digital Technological Change and the Cultural Industries." *Journal of Cultural Economics* 43, no. 2 (June 1, 2019): 189–210. <https://doi.org/10.1007/s10824-018-9336-2>.
- Tucker, Catherine, and Juanjuan Zhang. "How Does Popularity Information Affect Choices? A Field Experiment." *Management Science* 57, no. 5 (May 2011): 828–42. <https://doi.org/10.1287/mnsc.1110.1312>.
- Ursu, Raluca M. "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions." *Marketing Science* 37, no. 4 (August 2018): 530–52. <https://doi.org/10.1287/mksc.2017.1072>.
- Weitzman, Martin L. "Optimal Search for the Best Alternative." *Econometrica* 47, no. 3 (1979): 641–54. <https://doi.org/10.2307/1910412>.
- West, Patricia M., Dan Ariely, Steve Bellman, Eric Bradlow, Joel Huber, Eric Johnson, Barbara Kahn, John Little, and David Schkade. "Agents to the Rescue?" *Marketing Letters* 10, no. 3 (August 1, 1999): 285–300. <https://doi.org/10.1023/A:1008127022539>.