Information Design for Social Learning on a Recommendation Platform

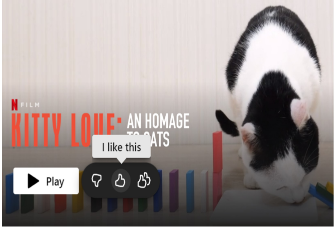
Chen Lyu
Peking University HSBC Business School

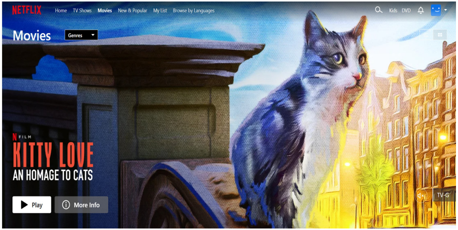
Aug 26, 2024
@ Rotterdam

Introduction

- Recommendation platforms are quite popular in daily life.
    - Goodreads for books
    - Netflix for movies
    - Yelp for restaurants
    - Tripadvisor for travel destinations

- To make better recommendations, a common practice of these platforms is to do "collaborative filtering".
- Platforms collect information generated from early consumers' experiences with a product, and use it to guide later consumers.





Introduction

- The recommendation policy plays a dual role:
  - Decide how past information is used
  - Decide whether new information will be generated.

  – this leads to a non-trivial dynamic information design problem.

- **Research question:** how a platform should design its recommendation policy for a new product in order to maximize the total consumer surplus generated on it. (Biased platform can also be handled in an extension.)

Introduction

- **The main consideration:** consumer incentives.
- Ideally, the platform should recommend trials for the new product as long as this is socially beneficial.
- However, because consumers do not internalize the value of information they generate, they may not want to follow such recommendations.
- The optimal design must choose when to recommend socially desirable but individually sub-optimal consumption efficiently, subject to that the consumers will be willing to follow.
- **A theme of the paper:** how this incentive problem should shape the platform's optimal design.

Model: Consumer Payoffs

- A product is of unknown quality $\tilde{\theta}$ taking values in $\{\theta_L, \theta_H\}$ ($\theta_L < 0 < \theta_H$). The platform and consumers share a common prior about it.
- It is launched at $t = 1$ and is available for $T < \infty$ periods.
- At each $t = 1, ..., T$, a short-lived consumer arrives at the platform and decides whether to consume the product ($a_t = 1$ if yes; $a_t = 0$ otherwise).
- The consumer's utility:

$$= \begin{cases} 0 & a_t = 0 \\ \tilde{\theta} & a_t = 1 \end{cases}$$

Model: The Learning Environment

- Whenever a consumer consumes the product, a signal about $\tilde{\theta}$ will be generated and privately observed by the platform.
  - Let s_i $(i \geq 1)$ denote the signal from the i'th consumption of the product.
  - Assume the signals are iid conditional on $\tilde{\theta}$.
- Before the product launches, the platform also receives a signal s_0 about $\tilde{\theta}$.
  - e.g., internal research or data about similar products.

Model: The Design Problem

- In each period, the platform can send a recommendation message to the current consumer. The consumer then makes her consumption decision.
- A *dynamic recommendation policy* decides what recommendation message to convey in each period based on any past signal realizations.
- The design problem: find such a policy, to which the platform can commit ex-ante, in order to maximize the total expected consumer surplus.
- By the revelation principle, it suffices to consider *incentive-compatible* policies with binary messages (i.e., to recommend or not).

Model: Comments

- An Important Model Feature: each consumer knows the product's launch time.
  - Many products have a public launch time (e.g., books, movies, podcasts, video games, etc).
  - Even if consumers only have partial information about the launch time, my design will be robust to the exact information they have.
- We must guarantee incentive compatibility separately for the consumer in each period.

Related Literature

- Main model of Kremer et al. (2014) considers fully revealing signals.
  - They focus on when to induce the first trial based on the platform's initial information.
- Che & Hörner (2018) considers Poisson learning with conclusive news.
  - They focus on a deterministic control problem over the recommendation intensity without news arrival.

Related Literature

- In comparison, my study considers general non-conclusive signals.
- My characterization of the optimal design is thus about whether to recommend the product in each period based on any current belief of the platform.
- This in particular allows me to interpret my results as regarding the optimal recommendation standard evolving over time, and do some new comparative statics.
- This necessitates a more general formulation of the design problem and a different approach to solving it.

Related Literature*

- Also related is a surging algorithm-oriented literature: an extension in Kremer et al. (2014), Papanastasiou et al. (2018), Mansour et al. (2020), etc.
  - Task: propose algorithms that can achieve better asymptotic performance as $T \to \infty$ (often measured by the decay rate of per-consumer regret).
  - Such performance criteria ignore welfare loss occurring within any finite time horizon, and can be insensitive to multiplicative increment in the welfare loss. (Note: $\frac{1}{T}$ and $\frac{2}{T}$ have the same decay rate in T.)
- My design may serve as a finite-horizon performance benchmark and help to inspire new algorithms with a non-asymptotic focus.

Solving the Optimal Design: The Constrained MDP

- Let p_t denote the platform's belief about $\tilde{\theta} = \theta_H$ given the information available at the beginning of time t.
- Process $(p_t)_{t=1}^{T}$ follows a Markov process controlled by the consumption decisions:

$$
\begin{aligned}
&p_1 \sim \mu_1 \\
&p_{t+1} | p_t, a_t \sim a_t \underbrace{G(\cdot | p_t)}_{\text{transition by Bayes updating}} + (1 - a_t) \underbrace{D(\cdot | p_t)}_{\text{Dirac measure}}
\end{aligned}
$$

- We can restrict to (randomized) Markov policy: $\phi := (\phi_t)_{t=1}^{T}$, where $\phi_t(p_t)$ is the recommendation probability given p_t at time t, which is also the probability for $a_t = 1$ when the consumer follows.

Solving the Optimal Design: The Constrained MDP

- The designer's problem:

$$\max_{\phi}\Big\{\sum_{t=1}^{T}\mathbb{E}_{\phi}[a_t u(p_t)]\Big\}$$

$$\text{s.t. } \mathbb{E}_{\phi}[a_t u(p_t)] \geq 0 \quad \forall t = 1, ..., T$$

$$p_t \text{ follows the process specified above}$$

where $u(p_t) := \theta_H p_t + \theta_L (1 - p_t)$

- Each IC constraint involves taking expectation over a_t and p_t at a particular time.

  – *this makes it a constrained Markov Decision Process*

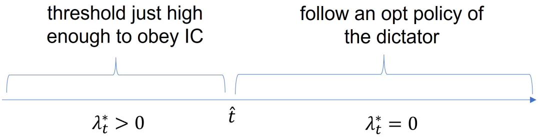
  – *the stochastic dynamic programming technique is not directly applicable.*

Solving the Optimal Design: The Constrained MDP

- I hence adopt a Lagrangian duality approach.
- I characterize the shadow values of the IC constraints, and then partially reduce the original problem into an unconstrained optimization over a Lagrangian function.
- This in the end allows me to fully solve the optimal design.
- To the best of my knowledge, this is the first paper that solves a constrained Markov decision process arising from a dynamic info design problem.

Solving the Optimal Design: The Optimal Policy

- The optimal design features *threshold policies*, and has a two-phase structure:



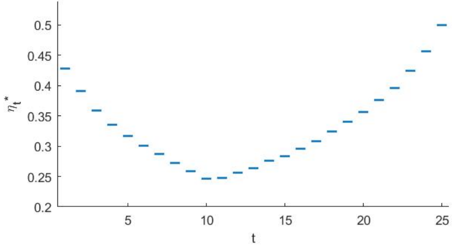
($\hat{t}$ can be pinned down as the first time when it is feasible to resume with dictator's optimal continuation policy.)

Implication I: Time Pattern of the Rec Standard

- The thresholds can be interpreted as time-varying rec standards.
- Optimal recommendation standard varies in a U-shaped pattern:

Proposition

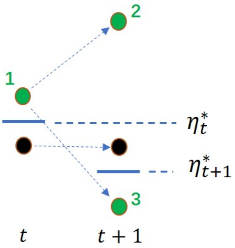
The thresholds $(\eta_t^*)_{t=1}^T$ of any optimal threshold policy satisfies: (a) $\eta_{t-1}^* > \eta_t^*$ for all $t \leq \widehat{t} - 1$; (b) $\eta_t^* < \eta_{t+1}^*$ for all $t \geq \widehat{t}$.



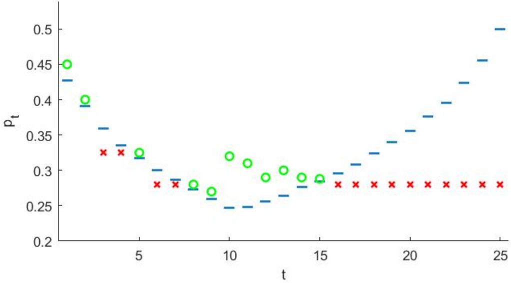
- Intuition: tension between the platform's desire to create information for later consumers and the need to fulfill the current consumer's IC constraint.

Implication I: Time Pattern of the Rec Standard*

- A more precise intuition about the decreasing part:



- An example path of recommendations under the optimal policy:



– The optimal recommendation can feature temporary suspensions following negative feedbacks when the product is young.

Implication III: CS w.r.t. Info Generation Rate

- How should the design be adjusted when consumption becomes more likely to yield informative signals (e.g., due to better feedback elicitation designs)?
- Ans: the recommendation standards should be lower for all t.

Proposition

Given any α, let $(\eta_t^*(\alpha))_{t=1}^T$ denote the thresholds of the optimal threshold policy. Then $\alpha_a < \alpha_b \implies \eta_t^*(\alpha_a) \geq \eta_t^*(\alpha_b)\,\forall t$, where the inequality is strict for all $t \in (1, T)$.

- Intuition: higher α implies:

  (1) higher informational value of consumption;

  (2) better information at any time, and thus consumers are more willing to follow the recommendations (ceteris paribus)

Implication IV: CS w.r.t. Consumer Arrival Rate*

- My main model has assumed that one consumer arrives for sure in each period.
- This is actually not needed – my framework easily accommodates random consumer arrives. All results carry over.
- CS: higher arrival rate $\implies$ lower recommendation standards

Proposition

Given any arrival rate ρ, let $(\eta_t^*(\rho))_{t=1}^T$ denote the thresholds of the optimal threshold policy. Then $\rho_a < \rho_b \implies \eta_t^*(\rho_a) \geq \eta_t^*(\rho_b)\, \forall t$, where the inequality is strict for all $t \in (1, T)$.

- Intuition: higher arrival rate implies
  (1) more consumers to come, and thus higher information value of consumption;
  (2) better information at any time, and thus consumers are more willing to follow the recommendations (ceteris paribus)

Extension: Biased Platform

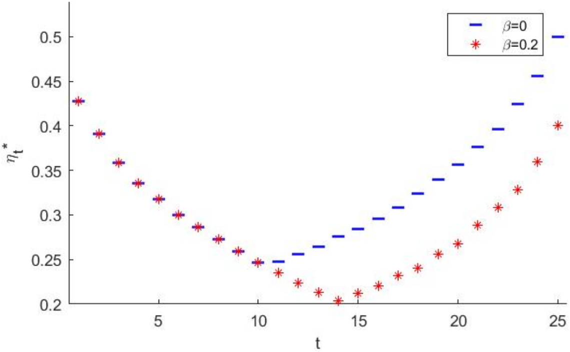
- My model can also incorporate biased platform who earn additional commission per consumption.
- The design problem can be re-written into:

$$\max_{\phi\in\Phi}\left\{\sum_{t=1}^{T}\mathbb{E}_{\phi}\left[a_t\left(u(p_t)+\beta\right)\right]\right\} \tag{1}$$

$$\text{s.t. } \mathbb{E}_{\phi}[a_t u(p_t)] \geq 0 \quad \forall t = 1, ..., T \tag{2}$$

where $\beta \geq 0$ measures the platform's bias (e.g., commission benefit).

- When β goes up, the following figure illustrates how the optimal design shifts.



Characterizing the Optimal Design: the Lagrangian Duality*

- The designer's problem:

$$\max_{\phi}\Big\{\sum_{t=1}^{T}\mathbb{E}_{\phi}[a_t u(p_t)]\Big\}$$
$$\text{s.t. } \mathbb{E}_{\phi}[a_t u(p_t)] \geq 0 \quad \forall t = 1, ..., T$$
$$p_t \text{ follows the process specified above}$$

To tackle this problem, I consider a Lagrangian duality approach.

- Given any Lagrangian multiplier $\lambda \in \mathbb{R}_{+}^{T}$, define the Lagrangian function for the designer's problem as:

$$\mathcal{L}(\phi; \lambda) = \sum_{t=1}^{T}\mathbb{E}_{\phi}[(1+\lambda_t)a_t u(p_t)] \tag{3}$$

Characterizing the Optimal Design: the Lagrangian Duality*

- Then, we have the strong duality result.

Lemma (duality)

Let w^* denote the optimal value of the designer's problem. Then,

$$w^* = \min_{\lambda \in \mathbb{R}_+^T} \sup_{\phi} \mathcal{L}(\phi; \lambda)$$

where the minimum is achieved by some non-negative λ^*. Given any such λ^*, a policy ϕ^* is optimal for the designer's problem if and only if:

(i) $\phi^* \in \arg\max_{\phi} \mathcal{L}(\phi; \lambda^*)$

(ii) $\lambda_t^* \mathbb{E}_{\phi^*}[a_t u(p_t)] = 0, \forall t = 1, ..., T$

(iii) $\mathbb{E}_{\phi^*}[a_t u(p_t)] \geq 0, \forall t = 1, ..., T$

Characterizing the Optimal Design: the Lagrangian Duality*

- The result implies that if we can find

$$\underbrace{\lambda^*}_{\text{shadow values}} \in \arg\min_{\lambda \in \mathbb{R}_+^T} \sup_{\phi} \mathcal{L}(\phi; \lambda) \quad \text{– dual problem}$$

  then the optimal design can be characterized by the optimization over the Lagrangian function, i.e.,

$$\phi^* \text{ is optimal} \implies \phi^* \in \arg\max_{\phi} \mathcal{L}(\phi; \lambda^*)$$

  – *an unconstrained problem.*

- Difficulty: the dual problem is also hard to solve.

- Now, the idea is to first extract some properties of λ^*, and see whether that will suffice for revealing certain features of the optimal design.
- Using the dual problem, I'm able to show:

Lemma (non-increasing shadow values)

There exists $\lambda^* \in \arg\min_{\lambda \in \mathbb{R}_+^T} \sup_\phi \mathcal{L}(\phi; \lambda)$ *such that* $\lambda_t^* \geq \lambda_{t+1}^* \, \forall t$.

– shadow values are non-increasing over time.

Characterizing the Optimal Design: Decreasing Multipliers*

- A rough intuition: As time passes

  1. more information accumulated $\Longrightarrow$ better informed choice $\Longrightarrow$ less sacrifice needed to obey IC;

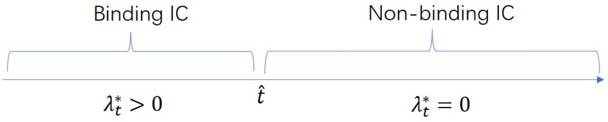
  2. shorter remaining time $\Longrightarrow$ lower informational value from consumption.

  – both suggest that relaxing later IC constraints is less helpful.

- (Proof: an inter-change argument.)

  $(\lambda_1, ..., \underbrace{\lambda_t, \lambda_{t+1}}_{<}, ...\lambda_T) \in \arg\min_{\lambda \in \mathbb{R}_+^T} \sup_\phi \mathcal{L}(\phi; \lambda)$

- A direct implication:



Characterizing the Optimal Design: Threshold Policies*

- The time pattern of λ_t^* also enables me to derive an important property of $\arg\max_\phi \mathcal{L}(\phi; \lambda^*)$

Lemma (non-increasing $\lambda_t \implies$ threshold solution)

If λ_t is non-increasing over t, then every solution to $\max_\phi \mathcal{L}(\phi; \lambda)$ is almost surely equivalent to a threshold policy.

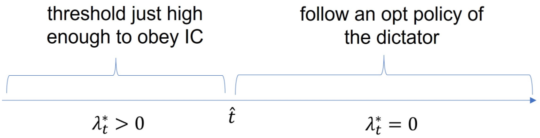
- This result requires properly weighing the dynamic and myopic values of consumption.

- In $\max_\phi \mathcal{L}(\phi; \lambda)$, when p_t increases
  - Myopic value of consumption increases
  - Dynamic information value of consumption may or may not

- With non-increasing $(\lambda_t)_{t=1}^T$, the change in the myopic value dominates. Thus the total value of consumption increases in p_t.

Characterizing the Optimal Design: The Optimal Policy*

- The previous lemmas together imply the following structure of optimal design:



($\hat{t}$ can be pinned down as the first time when it is feasible to resume with dictator's optimal continuation policy.)

- This enables an induction algorithm to construct an optimal policy ϕ^o (full characterization provided in the paper).

Summarizing Remarks

- I've studied how a platform can use its dynamic recommendations to direct consumers towards socially desirable information-generating consumption while maintaining their incentive in following the recommendations.
- I've shown that the optimal design generally features a "U"-shaped recommendation standard over a product's life.
- The optimal recommendation may involve temporary suspensions following negative consumer feedback.
- The optimal recommendation standards should be lowered when consumption becomes more informative or when consumers are arriving more frequently over time.

Summarizing Remarks

- Future research directions
  - Non-informational externality
  - Heterogeneous consumers with private information
  - Multiple products with unknown quality
  - Long-lived consumers who can wait
  - ...

Reference I

Che, Y.-K., & Hörner, J. (2018). Recommender systems as mechanisms for social learning. *The Quarterly Journal of Economics*, *133*(2), 871–925.

Kremer, I., Mansour, Y., & Perry, M. (2014). Implementing the "wisdom of the crowd". *Journal of Political Economy*, *122*(5), 988–1012.

Mansour, Y., Slivkins, A., & Syrgkanis, V. (2020). Bayesian incentive-compatible bandit exploration. *Operations Research*, *68*(4), 1132–1161.

Papanastasiou, Y., Bimpikis, K., & Savva, N. (2018). Crowdsourcing exploration. *Management Science*, *64*(4), 1727–1746.