Incentivizing Exploration by Heterogeneous Users

COLT 2018

Bangrui Chen, Peter Frazier

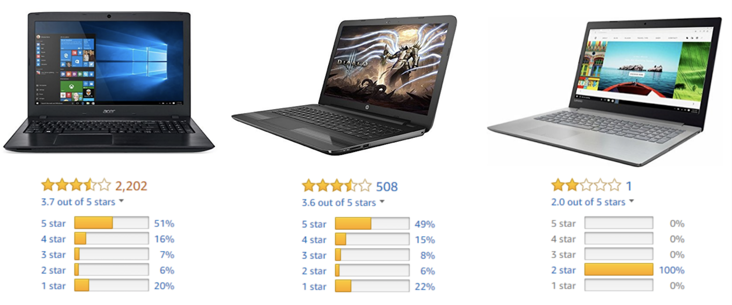
Cornell University
Operations Research and Information Engineering
bc496@cornell.com, pf98@cornell.edu

David Kempe

University of Southern California
Department of Computer Science
david.m.kempe@gmail.com

July 8, 2018

Customers Undervalue Exploration



- Incentives are misaligned:
  - Customers are myopic and want to **exploit**
  - Amazon wants customers to **explore**
- To fix this, Amazon can **incentivize exploration**

Previous Work

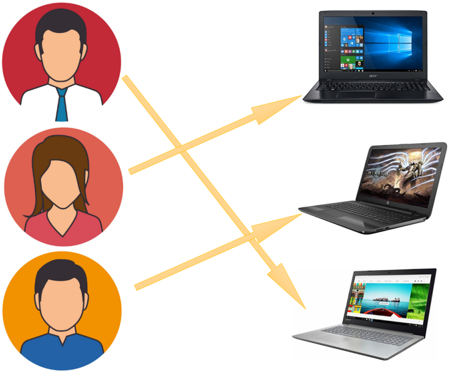
Without Money Transfer

- Kremer, Mansour & Perry 2014
- Mansour, Slivkins & Syrgkanis 2015
- Mansour, Slivkins, Syrgkanis & Wu 2016
- Mansour, Slivkins & Wu 2018
- Slivkins 2017

With Money Transfer

- Frazier, Kempe, Kleinberg & Kleinberg 2014
- Han, Kempe & Qiang 2015
- This paper

We Incentivize **Heterogeneous** Agents



- **Our setting**: Customers have different preferences
- **Challenge**: Amazon doesn't know these preferences
- **Opportunity**: Heterogeneity provides free explorations

Problem Setting

Agents

- Myopic agents arrive sequentially
- Agent t has linear utility with preference vector $\boldsymbol{\theta}_t \in \mathbb{R}^d$ drawn from known distribution F

Arms

- Each arm has an unknown feature vector $\boldsymbol{u}_i \in \mathbb{R}^d$
- Pulls give noisy observation of $\boldsymbol{u}_i$
- Everyone observes averages $\hat{\boldsymbol{u}}_{i,t}$ of each arm's past pulls

Agents' behavior

- Principal chooses payment $c_{t,i}$ for arm i at time t
- Agent t pulls arm $i_t = \arg\max_i \{\boldsymbol{\theta}_t \cdot \hat{\boldsymbol{u}}_{i,t} + c_{t,i}\}$

Principal's Goal

- Minimize cumulative regret with small cumulative payment

Algorithm Sketch

An arm is **payment-eligible** if:

- without incentives, its probability of being pulled is below a threshold
- AND it hasn't been pulled in a long-time

Our **algorithm**:

- If there is a payment-eligible arm, offer enough incentive to raise its probability of being pulled above the threshold
- Otherwise, let agents play myopically

Key Assumptions

- (**Every arm is someone's best**) Each arm is preferred by at least p fraction of users.
- (**Compact Support**) $\boldsymbol{\theta}$ has compact support.
- (**Few near-ties**) Let $q(z)$ be the proportion of agents with Utility(best arm) $\leq z +$ Utility(2^{nd} best arm).
  Then $q(z) \leq L \cdot z$ for all small enough z.

Main Result

Theorem 1

Our policy achieves:

- expected cumulative regret $O(Ne^{2/p} + LN \log^3(T))$,
- using expected cumulative payments of $O(N^2 e^{2/p})$.

We improve our bounds to polynomial in $1/p$ if

- agent preferences are discrete,
- OR we know a lower bound on p.

Discrete Preferences Give Constant Regret

Theorem 2

When agent preferences are discrete ($L = 0$):

- expected cumulative regret $O(N^2/p)$,
- using expected cumulative payments of $O(N/p)$.

- Regret and payment are constant in T
- The classical MAB has regret $O(\log T)$
- **Heterogeneity gives free exploration**

Questions?

Thanks for your time!