Incentivizing Exploration by Heterogeneous Users

COLT 2018

Bangrui Chen, Peter Frazier

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

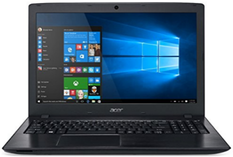
David Kempe

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

July 8, 2018

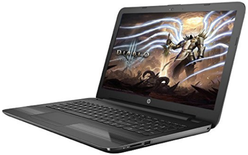
Motivation

Amazon wants users to ***explore***
Each customer only wants to buy one good item



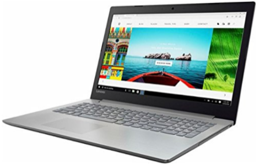
2,202
3.7 out of 5 stars

5 star	51%
4 star	16%
3 star	7%
2 star	6%
1 star	20%



508
3.6 out of 5 stars

5 star	49%
4 star	15%
3 star	8%
2 star	6%
1 star	22%



1
2.0 out of 5 stars

5 star	0%
4 star	0%
3 star	0%
2 star	100%
1 star	0%

Previous Work

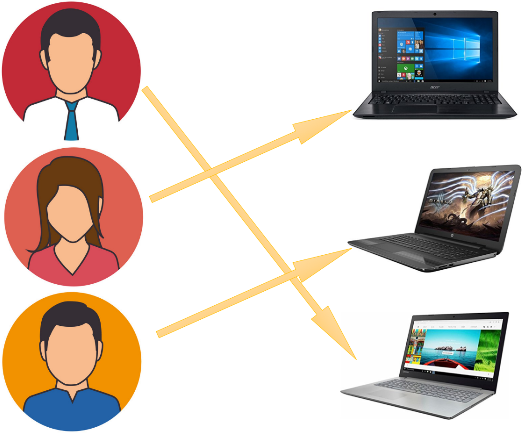
Without Money Transfer

- 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

Heterogeneity presents a new challenge



- **Customers prefer different kinds of items**
- **Amazon doesn't know which item each user prefers**

Heterogeneity Provides Free Exploration

- In the classical MAB: cumulative regret is $O(\log(T))$
- In incentizing exploration with heterogeneous users: we show, with assumptions, cumulative regret is $O(1)$
- Key insight: Heterogeneity provides free exploration
- Our contribution: First algorithm and analysis for incentivizing exploration when users have heterogeneous preferences over arms

Problem Setting

Agents

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

Arms

- Each arm has an unknown feature vector $\boldsymbol{u}_i \in \mathbb{R}^d$
- Agent t derives expected value $\boldsymbol{\theta}_t \cdot \boldsymbol{u}_i$ from pulling arm i
- Pulls gives 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 $i_t = \arg\max_i \{\boldsymbol{\theta}_t \cdot \hat{\boldsymbol{u}}_{i,t} + c_{t,i}\}$

The Principal's Goal

- Regret: $r_t = (\max_i \boldsymbol{\theta}_t \cdot \boldsymbol{u}_i) - \boldsymbol{\theta}_t \cdot \boldsymbol{u}_{i_t}$
- Payment: $c_t = c_{t,i_t}$
- **Principal's Goal**: Incentivize to minimize the cumulative regret while making a small cumulative payment

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 $O(N^2e^{2/p})$.

Special case: When agent preferences are discrete, i.e. $L = 0$, regret and payment are bounded by constants in T.

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

Questions?

Thanks for your time!