Privacy-Preserving Online Learning for Movie Recommendation

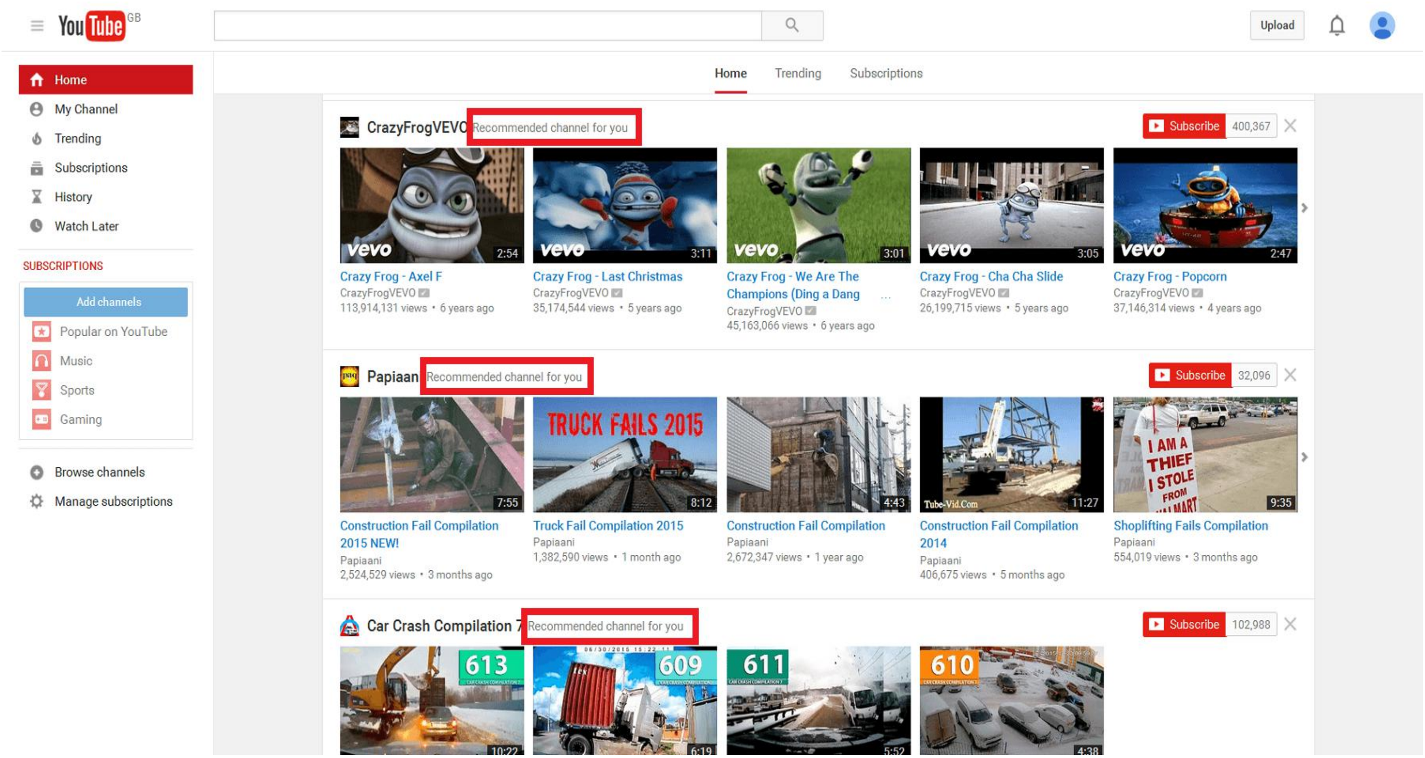
Yingxue Zhou

Dec. 4 2017

Outline

- Background Problem
- Online Learning
- Differential Privacy
- Algorithm
- Experiments

Background Problem: Movie Recommendations



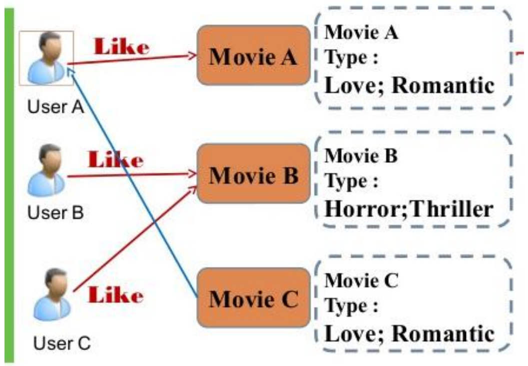
#YouTube

#personalized movie recommendation

#How to find the movie that matches a particular user’s preference?

#Privacy issue: Recommendation leaks user’s personal information.

Background Problem: Movie Recommendations



Different users with different preferences

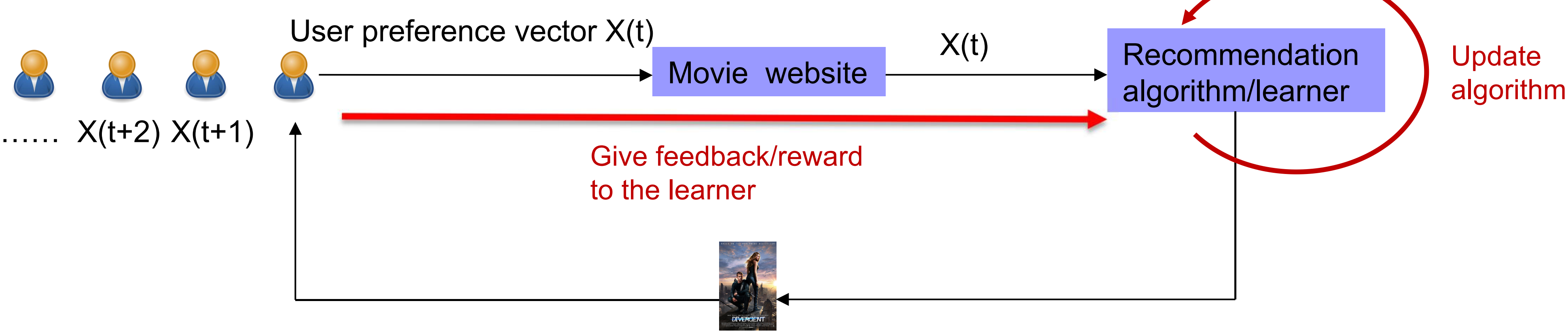
Movie types

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

Users are represented by row vectors

Static user database

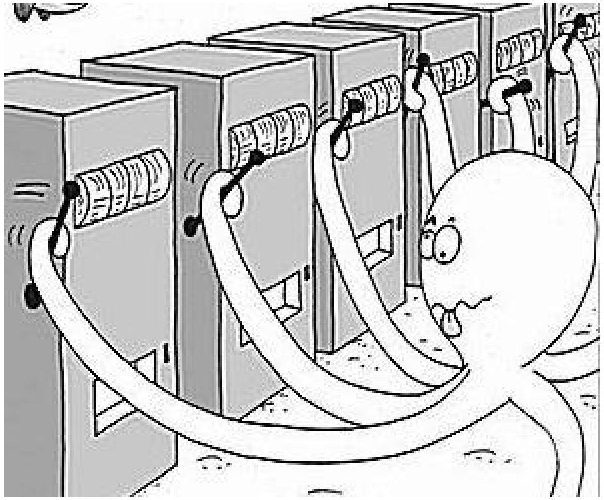
Actually, recommendation is a online learning task:



Goal: design a learning algorithm that can achieve recommendation and preserve user personal information.

Online Learning/Contextual Bandit

- **A gambler faces k slot-machines(arms).**
- **Each machine provides a random reward from unknown distribution specific to that machine.**
- **At each time slot, the gambler select one machine to play, and get a random reward.**
- **Goal: how to maximize the sum of rewards over all time slots.**



The stochastic bandit problem

Known parameters: number of arms K and (possibly) number of rounds $n \geq K$.
Unknown parameters: K probability distributions $\nu_1, \ldots, \nu_K$ on $[0, 1]$.

For each round $t = 1, 2, \ldots$

(1) the forecaster chooses $I_t \in \{1, \ldots, K\}$;
(2) given I_t, the environment draws the reward $X_{I_t,t} \sim \nu_{I_t}$ independently from the past and reveals it to the forecaster.

Online Learning/Contextual Bandit

The stochastic bandit problem

Known parameters: number of arms K and (possibly) number of rounds $n \geq K$.
Unknown parameters: K probability distributions $\nu_1, \ldots, \nu_K$ on $[0, 1]$.

For each round $t = 1, 2, \ldots$

(1) the forecaster chooses $I_t \in \{1, \ldots, K\}$;
(2) given I_t, the environment draws the reward $X_{I_t,t} \sim \nu_{I_t}$ independently from the past and reveals it to the forecaster.

For $i = 1, \ldots, K$ we denote by μ_i the mean of ν_i (mean reward of arm i). Let

$$\mu^* = \max_{i=1,\ldots,K} \mu_i \qquad \text{and} \qquad i^* \in \operatorname*{argmax}_{i=1,\ldots,K} \mu_i \ .$$

In the stochastic setting, it is easy to see that the pseudo-regret can be written as

$$\overline{R}_n = n\mu^* - \sum_{t=1}^{n} \mathbb{E}\left[\mu_{I_t}\right] \ .$$

Goal: Design a learning algorithm for the gambler to minimize the regret.

Online Learning/Contextual Bandit

For $i = 1, \ldots, K$ we denote by μ_i the mean of ν_i (mean reward of arm i). Let

$$\mu^* = \max_{i=1,\ldots,K} \mu_i \qquad \text{and} \qquad i^* \in \underset{i=1,\ldots,K}{\operatorname{argmax}} \, \mu_i \; .$$

In the stochastic setting, it is easy to see that the pseudo-regret can be written as

$$\overline{R}_n = n\mu^* - \sum_{t=1}^{n} \mathbb{E}\big[\mu_{I_t}\big] \; .$$

Formalize Recommendation as a bandit problem::

- At each time slot, recommender receives new user's contextual information.
- Choose a movie (arm) to recommend.
- Receive a random reward of recommended movie.
- Update strategy for next user.

Differential Privacy

Differential Privacy

X: The data *universe*.

$D \subset X$: The dataset (one element per person)

Definition: An algorithm M is ϵ-differentially private if for all pairs of neighboring datasets D, D', and for all outputs x:

$$\Pr[M(D) = x] \leq (1 + \epsilon) \Pr[M(D') = x]$$

Laplace Mechanism to achieve differential privacy:

M(D) = f(D)+Lap(b)

Laplace noise

directly query the database D

Algorithm

Part 1. Offline Estimation

1. Partition n users into m groups based on their contextual similarity.
2. Recommend all movies to them and gather rewards.
3. Compute average reward of different movie.

Part 2. Online Recommendation

At each time slot:

1. Receive new user, compute which group it belongs to.
2. Recommend the movie with highest average reward to the user.
3. Observe reward.
4. Add Laplace noise to this reward and update average reward.

Performance Metric:

Minimize the regret:

$$\overline{R}_n = n\mu^* - \sum_{t=1}^{n} \mathbb{E}\big[\mu_{I_t}\big] \ .$$

Experiments: Set Up

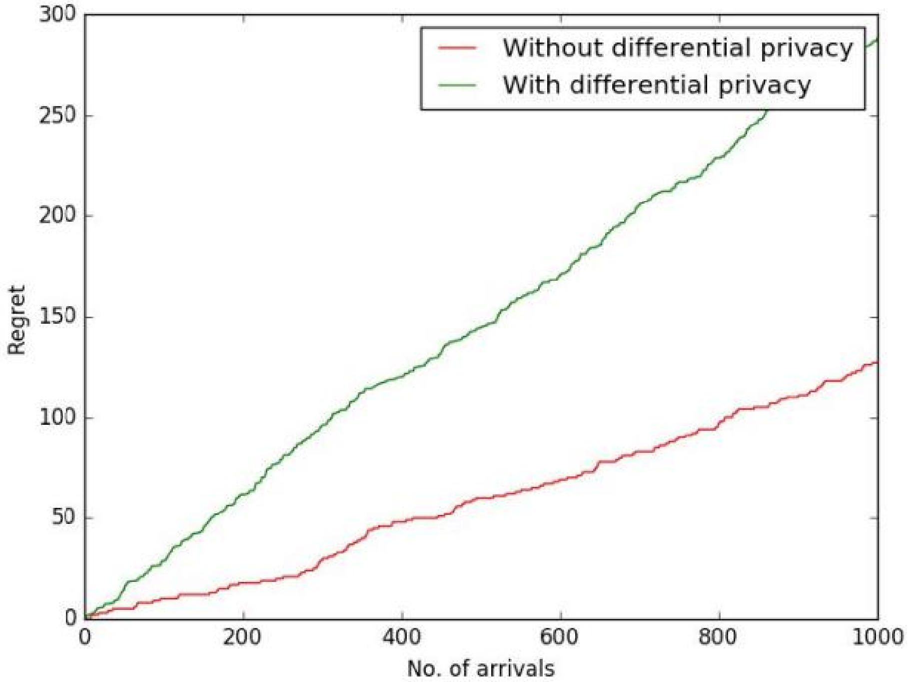
1. Dateset: the MovieLens dataset collected by the GroupLens Research Project at the University of Minnesota

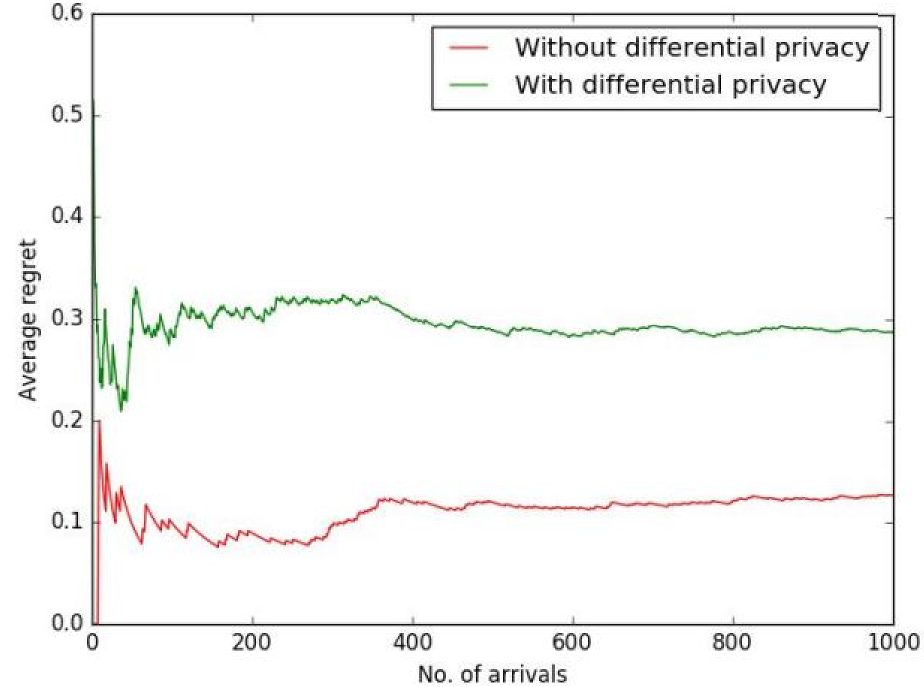
- 943 users
- 5 movie genres

2. Generate Bernoulli distribution to simulate user's reward/feedback.

3.Plot the regret function to show the performance of the proposed algorithm.

Experiments: Results





References

1. Dwork, Cynthia, et al. "Calibrating noise to sensitivity in private data analysis." Theory of Cryptography Conference. Springer BerlinHeidelberg, 2006.
2. Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." Foundations and Trends in Theoretical ComputerScience 9.3-4 (2014): 211-407.
3. Li L, Chu W, Langford J, et al. A contextual-bandit approach to personalized news article recommendation[C]. Proceedings of the 19th international conference on World wide web. ACM, 2010: 661-670.
4. Auer P, Cesa-Bianchi N, Fischer P. Finite-time analysis of the multiarmed bandit problem[J]. Machine learning, 2002, 47(2-3): 235-256.
5. Dwork C. Differential privacy: A survey of results[C]. International Conference on Theory and Applications of Models of Computation.Springer Berlin Heidelberg, 2008: 1-19.
6. Jeckmans A J P, Beye M, Erkin Z, et al. Privacy in recommender systems[M]. Social media retrieval. Springer London, 2013: 263-281.
7. F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transaction on Intellligent System(TiiS) 5, 4, Article 19.
8. Mishra N, Thakurta A. Private stochastic multi-arm bandits: From theory to practice[C]. ICML Workshop on Learning, Security, and Privacy.2014

Thank You!