Causal Inference for Recommendation

Dawen Liang¹, Laurent Charlin², David Blei¹

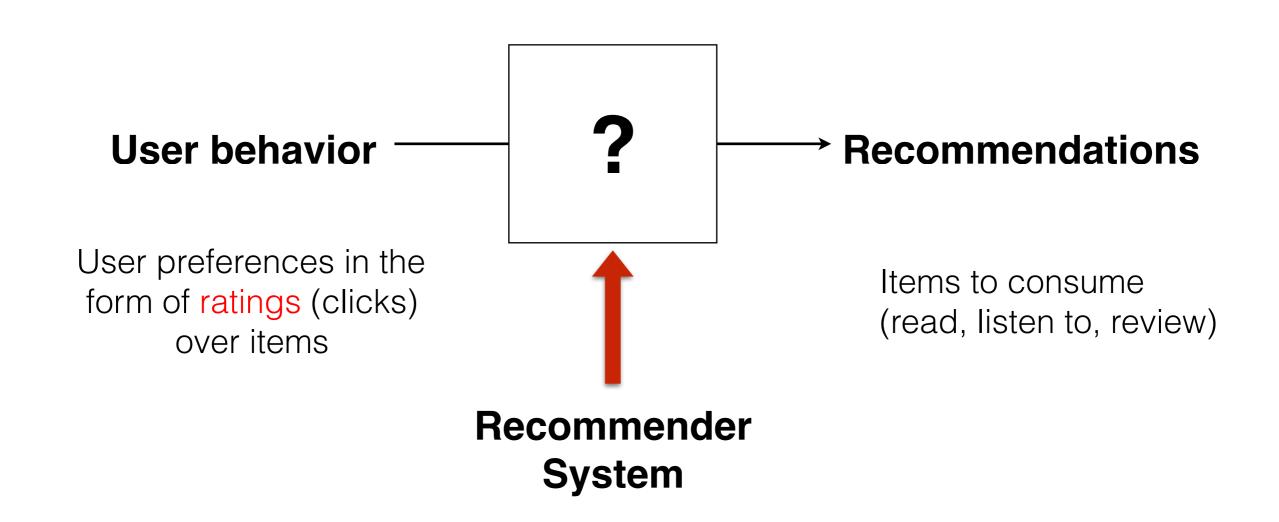
¹Columbia University

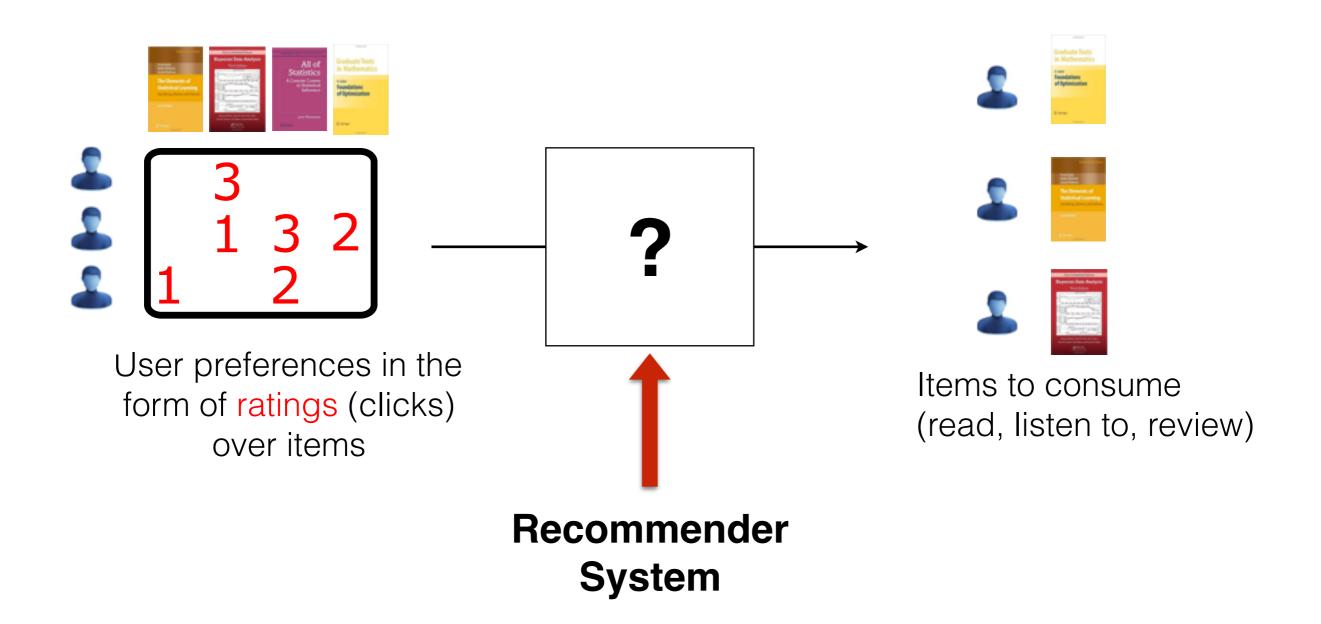
²HEC Montréal



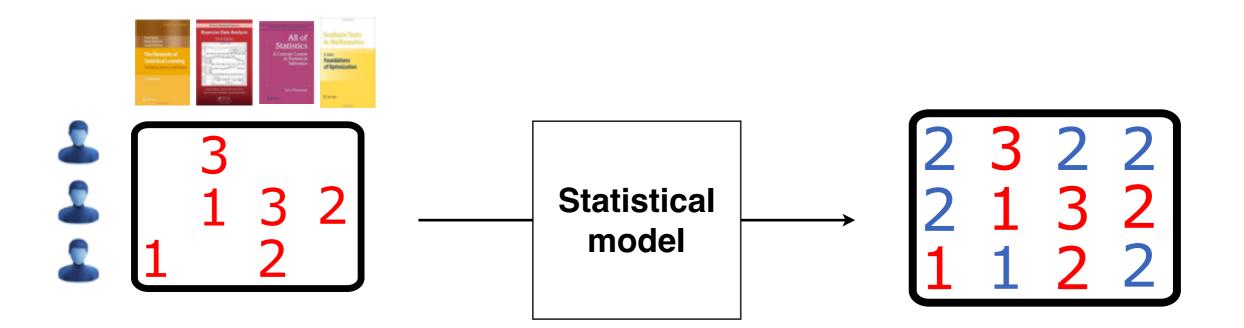




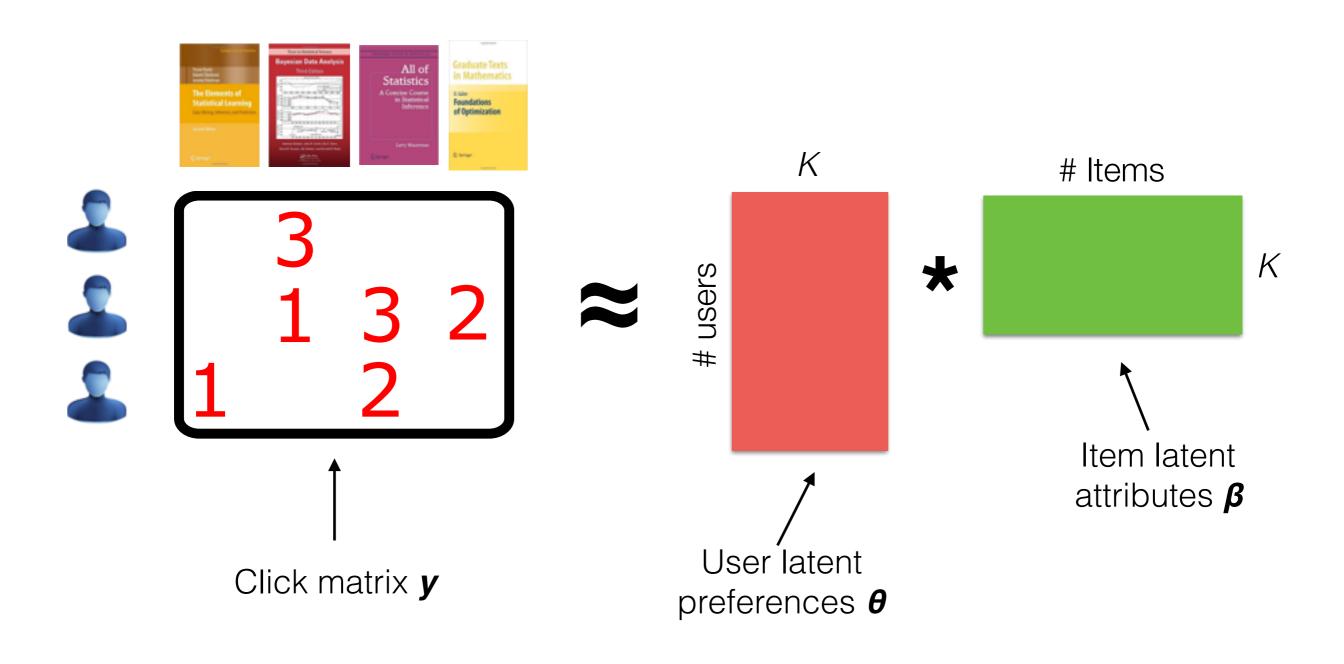


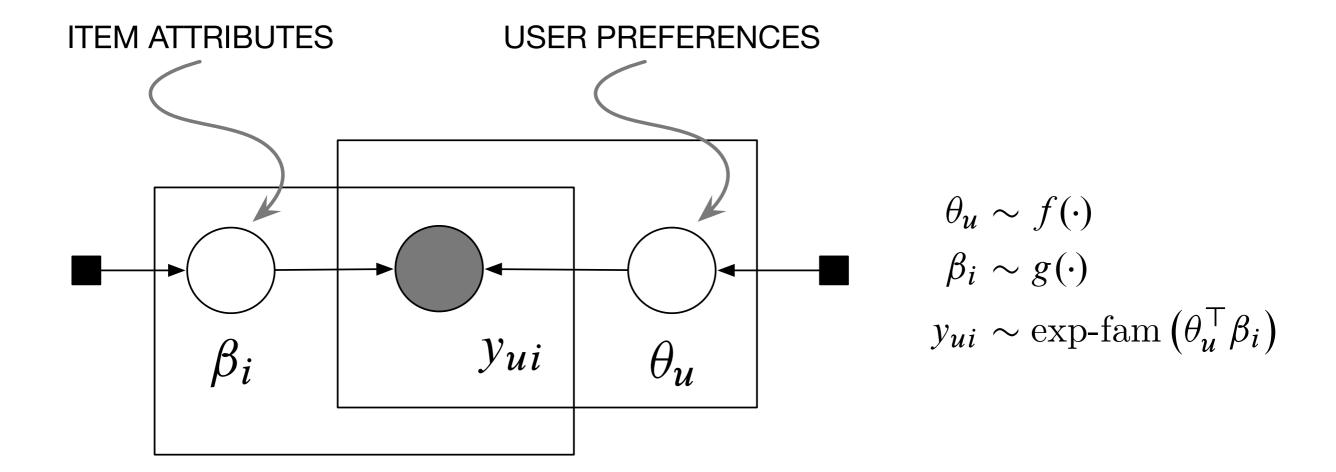


Recommendation as matrix completion



Classical solution: matrix factorization





- User u is associated with latent preference θ_u
- Item i is associated with latent attributes βi
- Whether a user clicks on an item comes from a exponential family distribution, e.g., Poisson, Bernoulli, Gaussian

- Classical Bayesian inference:
 - Condition on click data \mathbf{y} to estimate or approximate the posterior $p(\mathbf{\theta}, \mathbf{\beta} \mid \mathbf{y})$
 - Obtain estimates of user preferences and item attributes
- Form prediction with posterior predictive distribution
- Recommend based on the ranking of the predictions

"All Things Airplane"

Flying Solo Crew-Only 787 Flight Is Approved By FAA All Aboard Rescued After Plane Skids Into Water at Bali Airport Investigators Begin to Test Other Parts On the 787 American and US Airways May Announce a Merger This Week

"Personal Finance"

In Hard Economy for All Ages Older Isn't Better It's Brutal Younger Generations Lag Parents in Wealth-Building Fast-Growing Brokerage Firm Often Tangles With Regulators The Five Stages of Retirement Planning Angst Signs That It's Time for a New Broker

Example components from New York Times click data

How can we connect recommendation with causal inference?

Causal inference

Expose a unit to a treatment

"What would have happened if patient A received the treatment X?"

Biased data from observational studies

Recommendation

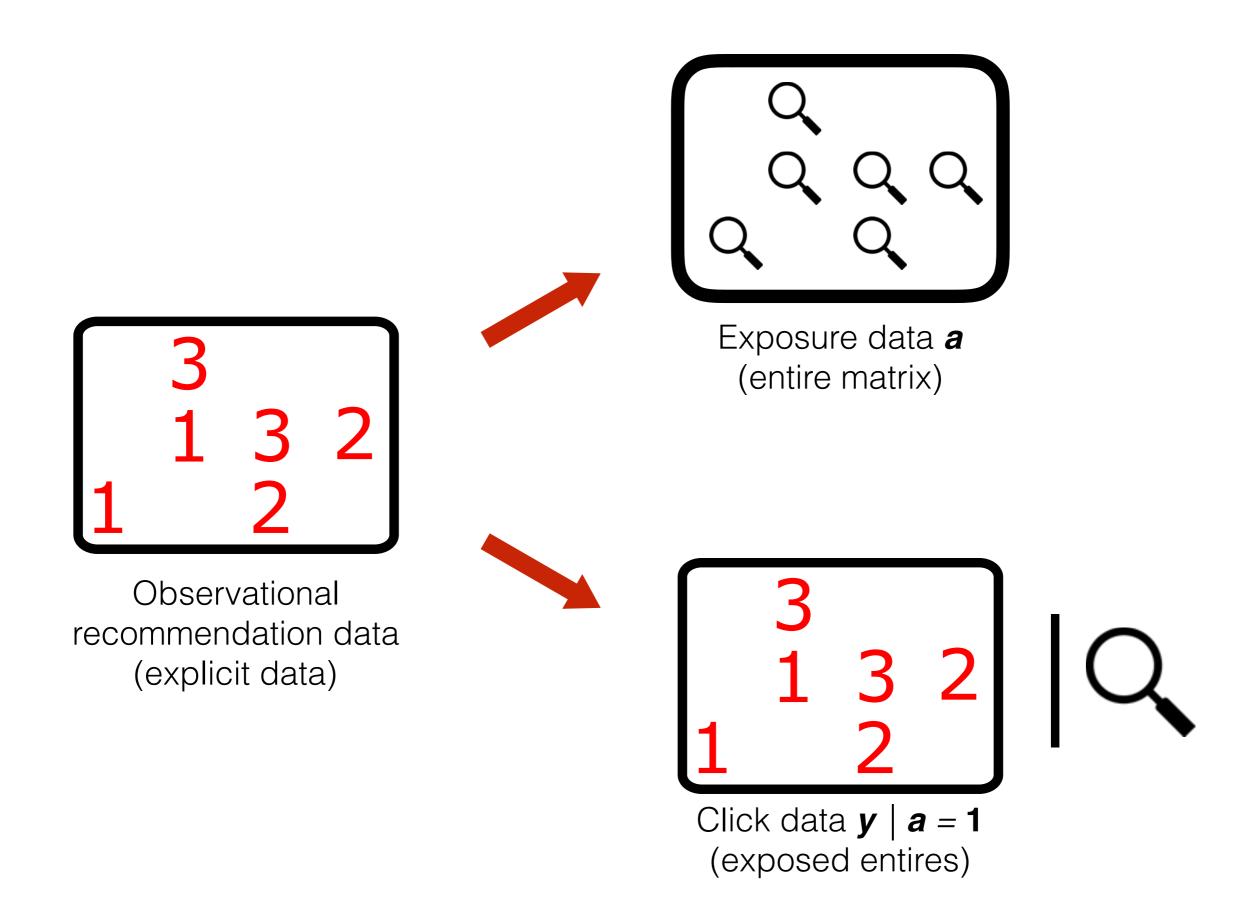
Expose a user to an item

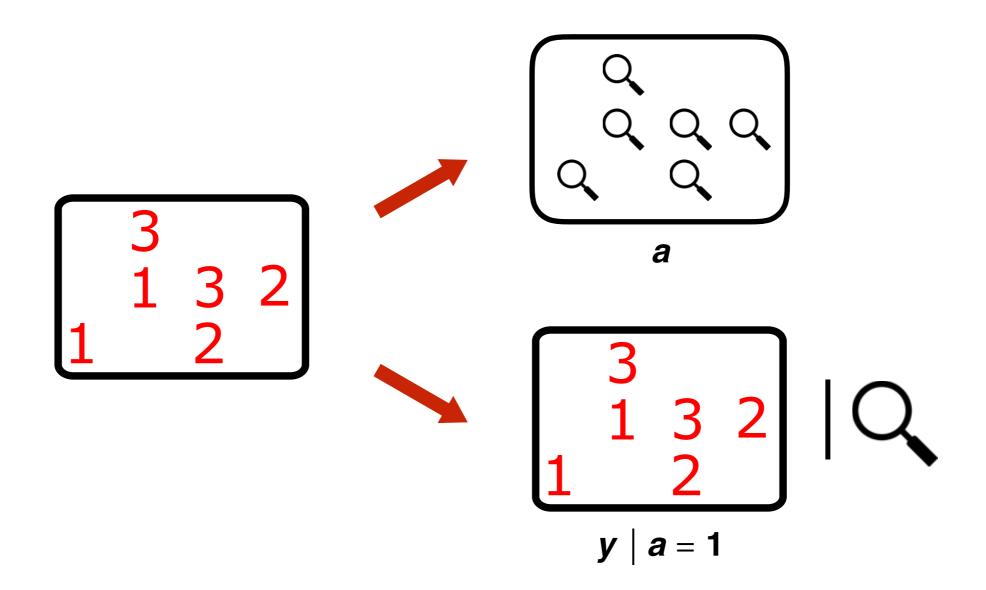
"What would have happened if user A was recommended item X?"

Biased data from logged user behavior

<u>'?</u>

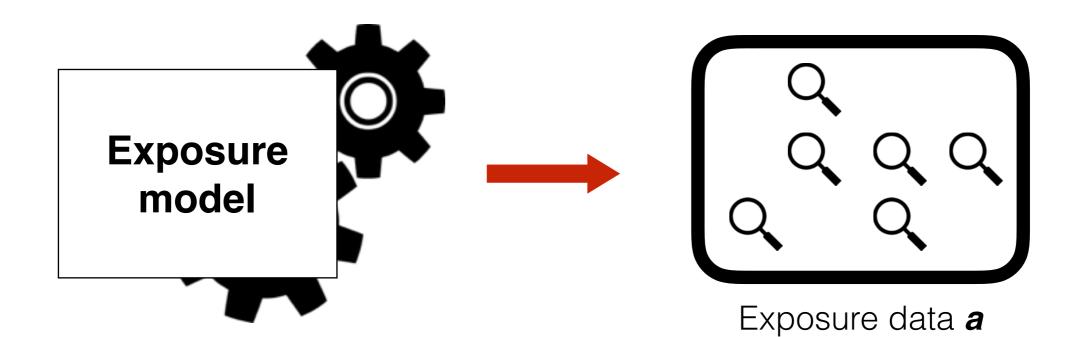
7

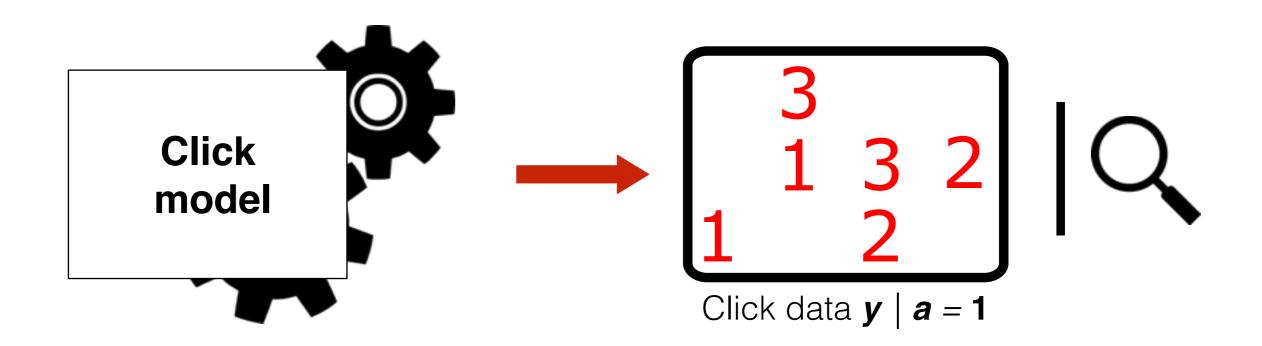




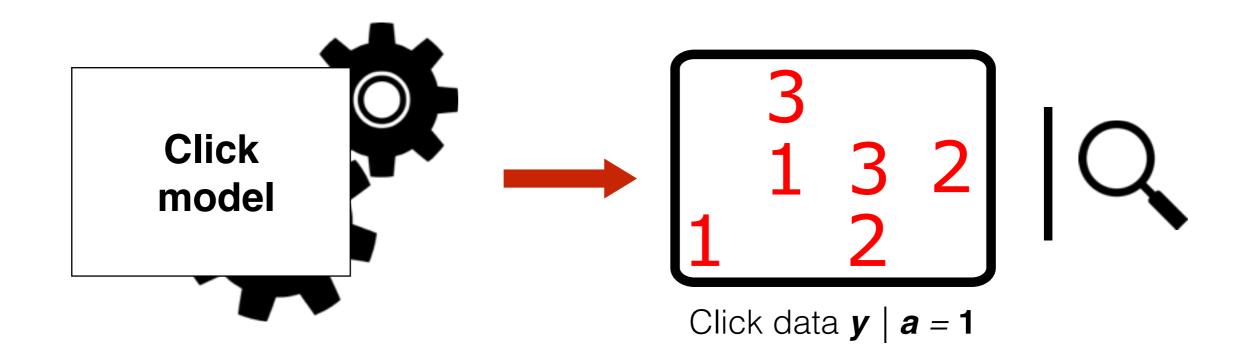
Generative process from a *user exposure* perspective

- Each user is exposed to certain amount of items
- Among these items, the user decides to click on (like) some of the items, and skip (dislike) some others



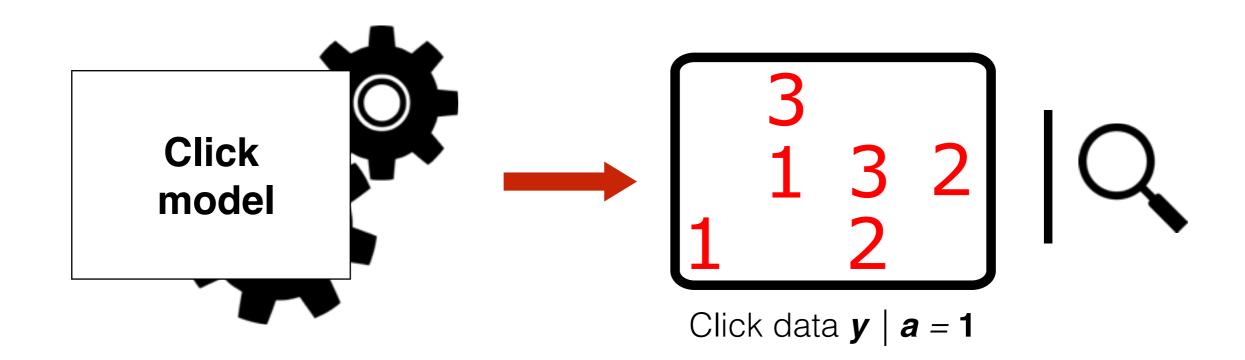


One problem with classical Bayesian inference



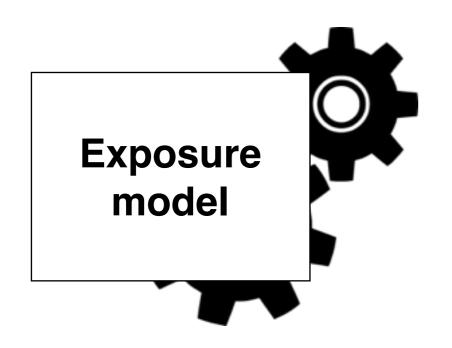
- Data is biased by the exposure data
 - Items that are more exposed have too much influence
- Inferring the click model from click data alone leads to a biased model

One problem with classical Bayesian inference



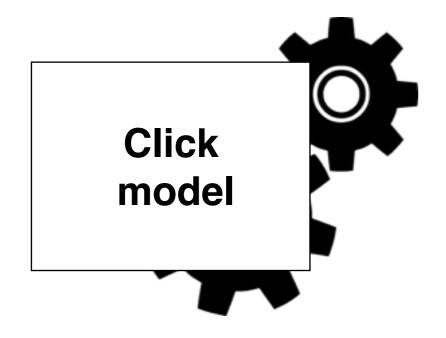
- Remedy A: Run a randomized experiment where we control exposure
 - Hard to do in practice
- Remedy B: Re-weight the data to make it appear as if it came from a randomized experiment
 - Weight each observation by the inverse probability of being exposed (inverse propensity weighting)

Joint model of exposure and click



Popularity exposure

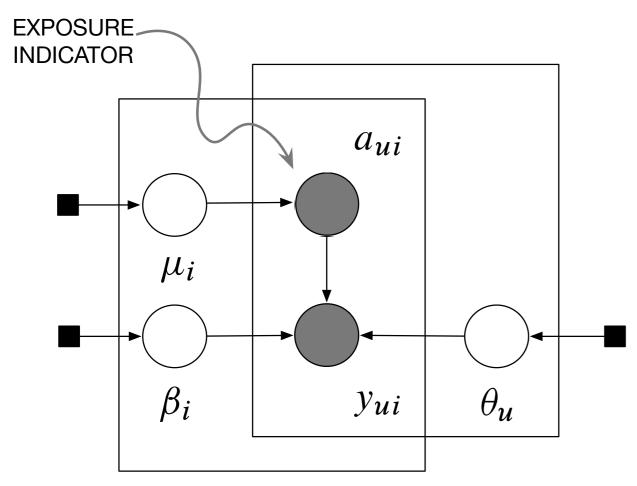
 $a_{ui} \sim \text{Bernoulli}(\mu_i)$



$$y_{ui} \mid a_{ui} = 0 \sim \delta_0$$

 $y_{ui} \mid a_{ui} = 1 \sim \exp\text{-fam}(\theta_u^{\top} \beta_i)$

Joint model of exposure and click



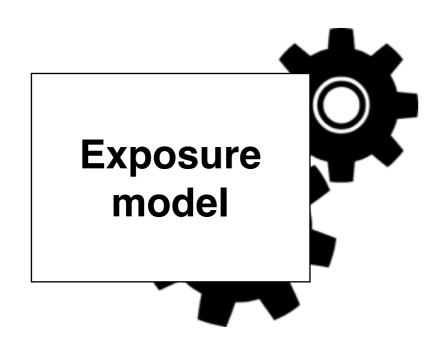
Popularity exposure

 $a_{ui} \sim \text{Bernoulli}(\mu_i)$

$$y_{ui} \mid a_{ui} = 0 \sim \delta_0$$

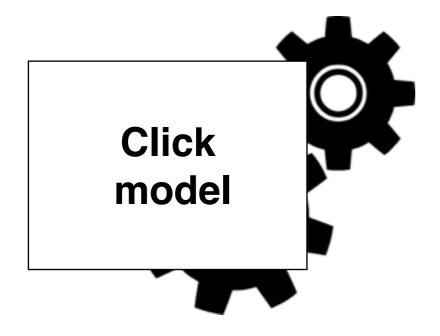
 $y_{ui} \mid a_{ui} = 1 \sim \exp\text{-fam}(\theta_u^{\top} \beta_i)$

Joint model of exposure and click



Poisson factorization exposure

$$a_{ui} \sim \text{Poisson}(\pi_u^{\top} \gamma_i)$$



$$y_{ui} \mid a_{ui} = 0 \sim \delta_0$$

 $y_{ui} \mid a_{ui} = 1 \sim \exp\text{-fam}(\theta_u^{\top} \beta_i)$

How is this different from standard causal inference?

Causal inference

Single causal question

Many observations to estimate the causal effect

Recommendation

Many causal questions (one per user-item pair)

Causal outcomes are related (governed by the same set of parameters)

The algorithm

- Fit the exposure model (popularity or Poisson factorization)
- Use exposure model predictions to re-weight the click data
 - It down-weights the influence of more exposed items and up-weights the influence of less exposed items
- Use the causally-inferred click model for predictions

Forming predictions

We study two ways of predicting

Conditional prediction

$$\mathbb{E}[y_{ui} \mid a_{ui} = 1, \mathcal{D}] = \theta_u^{\top} \beta_i$$

Marginal prediction

$$\mathbb{E}[y_{ui} \mid \mathcal{D}] = \mathbb{P}(a_{ui} = 1 \mid \pi_{ui}, \mathcal{D}) \cdot \mathbb{E}[y_{ui} \mid a_{ui} = 1, \mathcal{D}]$$

$$\uparrow$$

$$\text{how likely user } u \text{ user } u\text{'s preference will seek out item } i \text{ towards item } i$$

Why might this work?

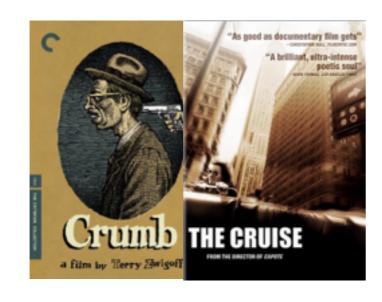
Drama lover



Regular recommender system



More dramas





Why might this work?

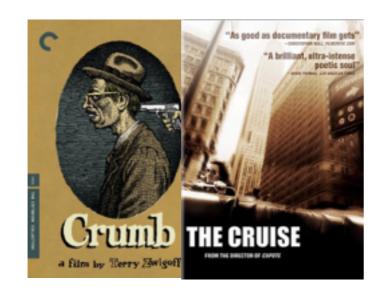
Drama lover



Causal recommender system



More dramas







Related work

Self-selection process

[Marlin & Zemel'09, Schnabel et al.'16]

Multi-arm bandits

[Li et al.'15, Li et al.'10, Vanchinathan et al.'14, Zhao et al.'13]

Empirical study

| | ML-1M | ML-10M | Yahoo-R3 | ArXiv |
|----------------|-------|--------|----------|--------|
| # of users | 6,040 | 69,878 | 15,400 | 26,541 |
| # of items | 3,706 | 10,677 | 1,000 | 80,082 |
| # of exposures | 1.0M | 10.0M | 0.3M | 1.9M |
| % of exposures | 4.47% | 1.34% | 2.02% | 0.09% |

- ML-1M, ML-10M, Yahoo-R3:
 - $a_{ui} = 1$: user *u* having rated item *i*
- ArXiv:
 - $a_{ui} = 1$: user u having viewed the abstract of paper i.
 - $y_{ui} = 1$ if she downloaded the paper and $y_{ui} = 0$ otherwise

- Exposure model: Popularity or Poisson factorization
- Model fitting: Train the click model causally with inverse propensity weighting, or observationally with classical inference.
- Prediction: Conditional prediction or marginal prediction

Among these methods are two baselines from previous work:

- "Probabilistic matrix factorization", Salakhutdinov & Mnih, NIPS 2008
- "Recommendations as treatments: Debiasing learning and evaluation", Schnabel et al., ICML 2016

| | ML-1M | ML-10M | Yahoo-R3 | ArXiv |
|-----|-------|--------|----------|-------|
| PMF | 13.0% | 5.4% | 15.1% | 18.4% |

Mean average rank: on average at what percentile a heldout item will be ranked (smaller is better)

- Test data is from the joint model of exposure and click
- Skewed test data shows even better performance from causally trained model
 - Approximate an intervention on the exposure

Current and future work

- Causality with hidden exposure (implicit data)
- "Proper" Bayesian inference