Unbiased Learning to Rank with Biased Feedback

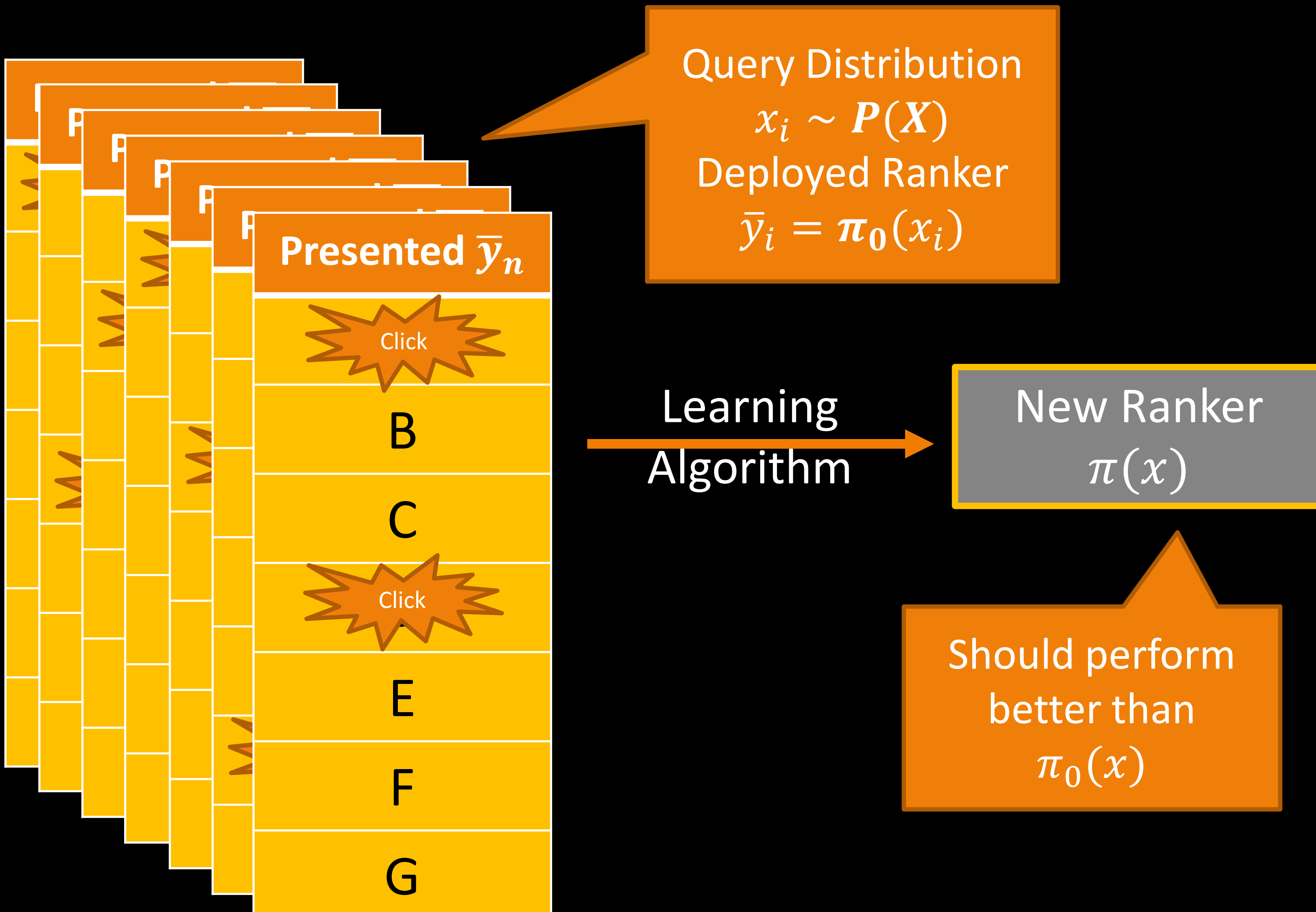
CS7792 Counterfactual Machine Learning – Fall 2018

Thorsten Joachims

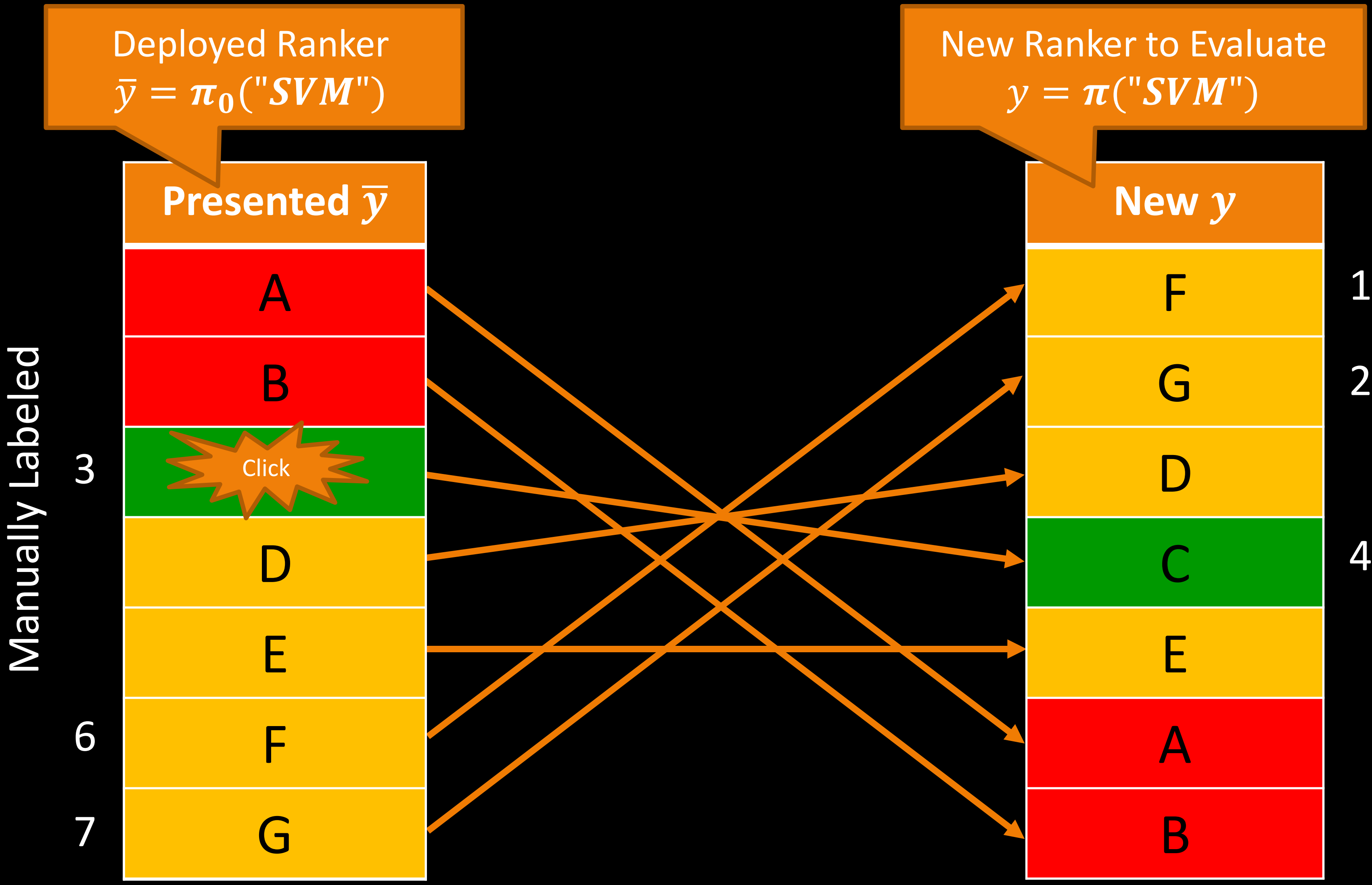
Departments of Computer Science and Information Science
Cornell University

- T. Joachims, A. Swaminathan, T. Schnabel, Unbiased Learning-to-Rank with Biased Feedback, International Conference on Web Search and Data Mining (WSDM), 2017.

Learning-to-Rank from Clicks



Evaluating Rankings



Evaluation with Missing Judgments

- Loss: $\Delta(y|r)$
  - Relevance labels $r_i \in \{0,1\}$
  - This talk: rank of relevant documents

$$\Delta(y|r) = \sum_i rank(i|y) \cdot r_i$$

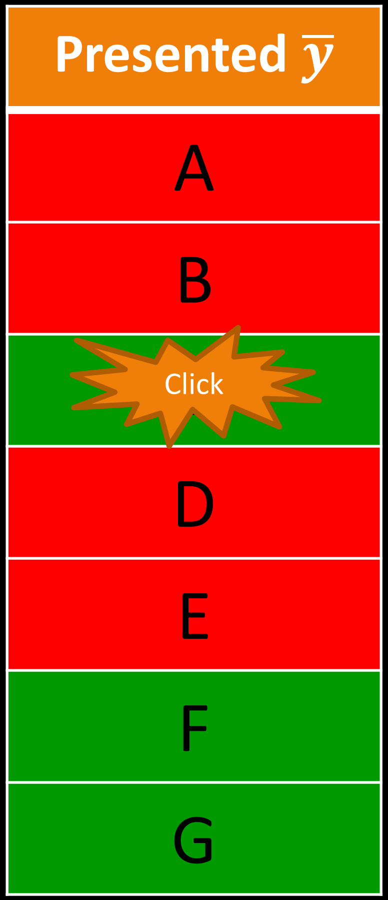
- Assume:
  - Click implies observed and relevant:

$$(c_i = 1) \leftrightarrow (o_i = 1) \wedge (r_i = 1)$$

- Problem:
  - No click can mean not relevant OR not observed

$$(c_i = 0) \leftrightarrow (o_i = 0) \vee (r_i = 0)$$

→ Understand observation mechanism



Inverse Propensity Score Estimator

- Observation Propensities $Q(o_i = 1|x, \bar{y}, r)$
  - Random variable $o_i \in \{0,1\}$ indicates whether relevance label r_i for is observed
- Inverse Propensity Score (IPS) Estimator:

$$\hat{\Delta}(y|r,o) = \sum_{i:o_i=1} \frac{rank(i|y) \cdot r_i}{Q(o_i = 1|\bar{y}, r)}$$

$$= \sum_{i:o_i=1 \wedge r_i=1} \frac{rank(i|y)}{Q(o_i = 1|\bar{y}, r)}$$

$$= \sum_{i:c_i=1} \frac{rank(i|y)}{Q(o_i = 1|\bar{y}, r)}$$

New Ranking

Need to know the propensities only for relevant/clicked docs.

- Unbiasedness: $E_o\left[\hat{\Delta}(y \mid r, o)\right] = \Delta(y|r)$

Presented $\bar{y}$	Q
A	1.0
B	0.8
C	0.5
D	0.2
E	0.2
F	0.2
G	0.1

[Horvitz & Thompson, 1952] [Rubin, 1983] [Zadrozny et al., 2003] [Langford, Li, 2009] [Swaminathan & Joachims, 2015]

ERM for Partial-Information LTR

- Unbiased Empirical Risk:

$$\hat{R}_{IPS}(\pi) = \frac{1}{N} \sum_{(x,\bar{y},c)\in S} \sum_{i:c_i=1} \frac{rank(i|\pi(x))}{Q(o_i = 1|\bar{y}, r)}$$

Consistent Estimator of True Performance

- ERM Learning:

$$\hat{\pi} = \underset{\pi\in\Pi}{\operatorname{argmin}}\left[\hat{R}_{IPS}(\pi)\right]$$

Consistent ERM Learning

- Questions:
  - How do we optimize this empirical risk in a practical learning algorithm?
  - How do we define and estimate the propensity model $Q(o_i = 1|\bar{y}, r)$? → Next week by Aman

BLBF vs. LTR

Batch Learning from Bandit Feedback	Learning to Rank from Implicit Feedback
• Atomic actions	• Combinatorial actions
• Action y chosen by π_0 influences feedback	• Action y chosen by π_0 influences feedback
• Observe loss $\delta(x, y)$ for action y chosen by π_0.	• Observe partial information about loss $\delta(x, y)$ for multiple y
• Interventional → Logged propensities	• Interventional + Observational (user)

Propensity-Weighted SVM Rank

- Data: $$S = \left(x_j, d_j, D_j, q_j\right)^n$$

Query | Clicked | Others | Propensity

- Training QP:

$$w^* = \underset{w,\xi \geq 0}{\operatorname{argmin}} \frac{1}{2} w \cdot w + \frac{C}{n} \sum_j \frac{1}{q_j} \sum_i \xi_j^i$$
$$\forall \bar{d}^i \in D_1 : w \cdot \left[\phi(x_1, d_1) - \phi\left(x_1, \bar{d}^i\right)\right] \geq 1 - \xi_1^i$$
$$\vdots$$
$$\forall \bar{d}^i \in D_n : w \cdot \left[\phi(x_n, d_n) - \phi\left(x_n, \bar{d}^i\right)\right] \geq 1 - \xi_n^i$$

Optimizes convex upper bound on unbiased IPS risk estimate!

- Loss Bound:

$$\forall w : rank(d, sort(w \cdot \phi(x, d)) \leq \sum_i \xi^i + 1$$

[Joachims et al., 2002]

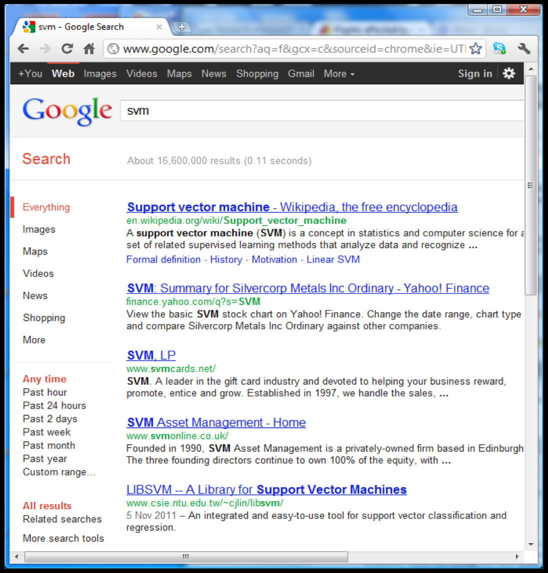
Position-Based Propensity Model

- Model:

$$P\big(c_i = 1 | r_i, rank(i|\bar{y})\big) =$$
$$P(o_i = 1 | rank(i|\bar{y}))$$
$$\cdot P(c_i = 1 | r_i, o_i = 1)$$

Propensity
$Q(o_i = 1 | x, \bar{y}, r)$

- Assumptions
  - Examination only depends on rank
    → $Q(o_i = 1 | rank(i|\bar{y})) = q_r$
  - Clicks reveal relevance if examined
    $P(c_i = 1 | r_i = 1, o_i = 1) = 1$
    and
    $P(c_i = 1 | r_i, o_i) = 0$ otherwise



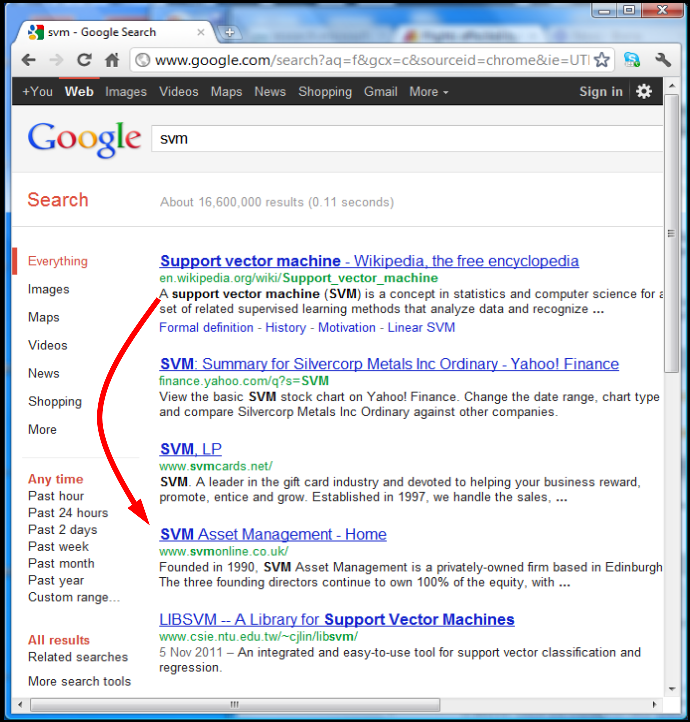
Estimating the Propensities

- Experiment:
  - Click rate at rank 1:

    $$q_1 \cdot E(r_i = 1 | rank(i|\bar{y}) = 1)$$

- Intervention:
  - swap results at rank 1 and rank k
  - Click rate at rank k:

    $$q_k \cdot E(r_i = 1 | rank(i|\bar{y}) = 1)$$



$$\rightarrow \frac{q_1}{q_k} = \frac{Click\ rate\ at\ rank\ 1}{Click\ rate\ at\ rank\ k\ after\ swap}$$

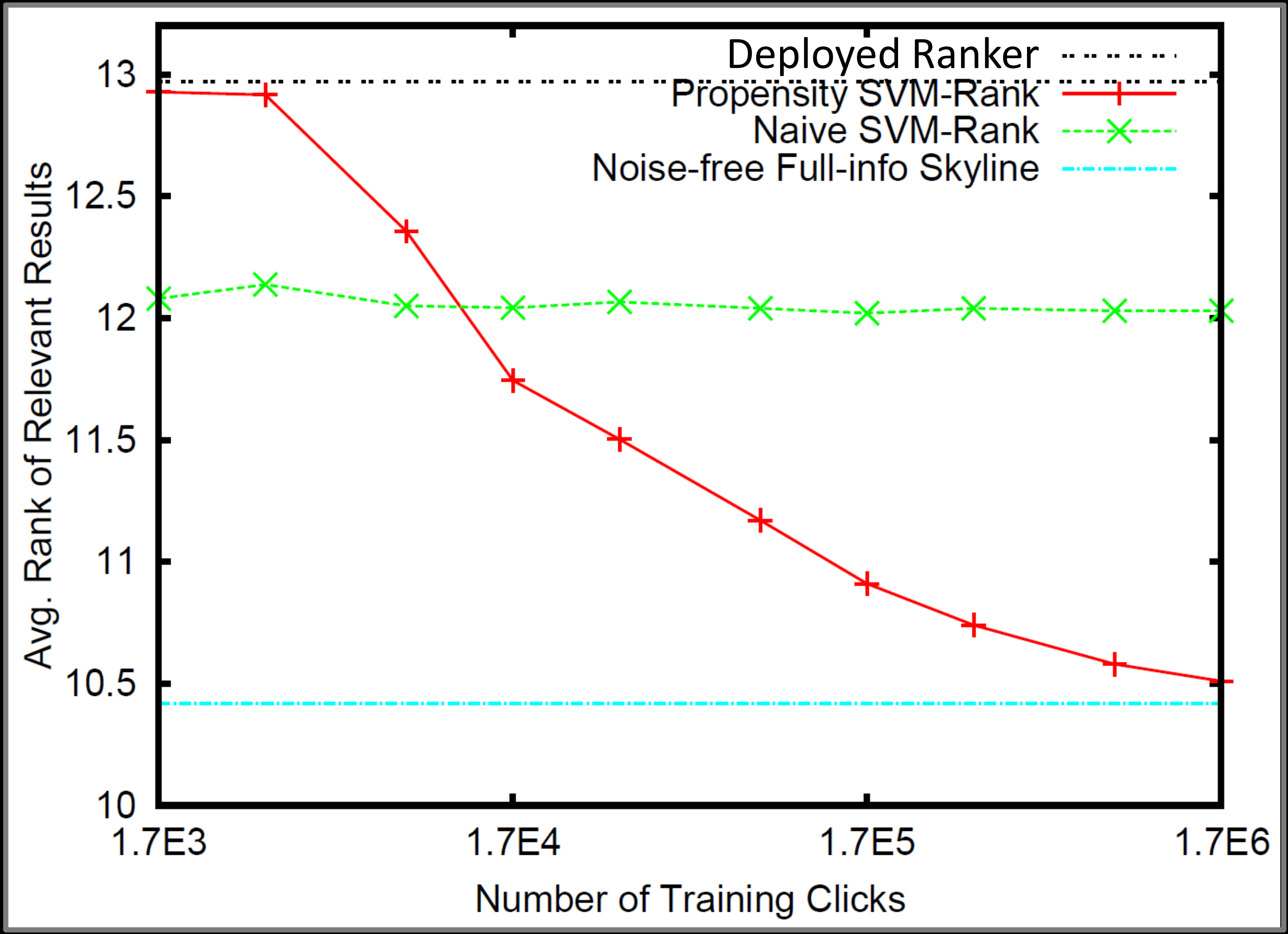
[Langford et al., 2009; Wang et al., 2016]

Experiments

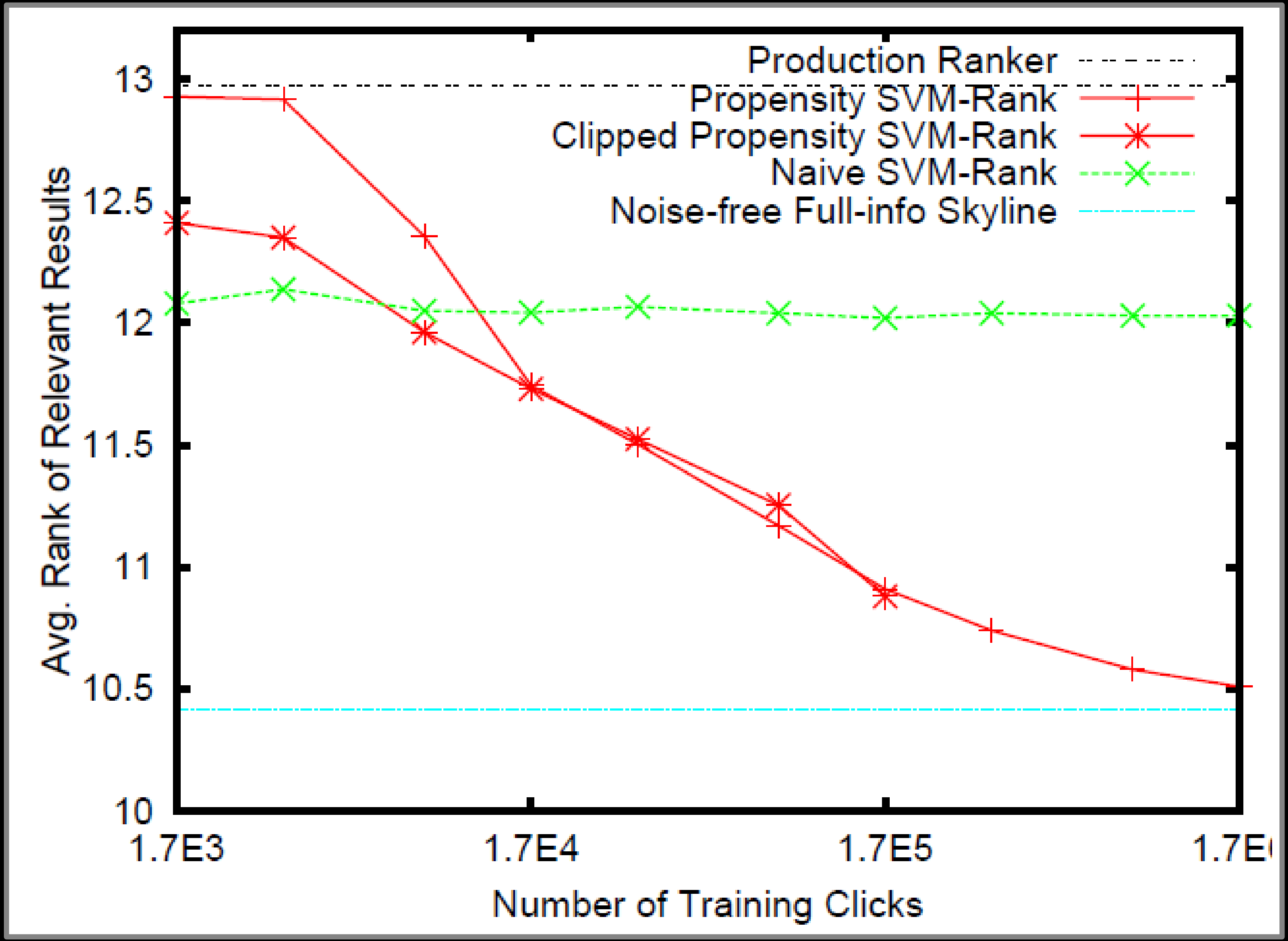
- Yahoo Web Search Dataset
  - Full-information dataset
  - Binarized relevance labels
- Generate synthetic click data based on
  - Position-based propensity model with $q_r = \left(\frac{1}{r}\right)^{\eta}$
  - Baseline "deployed" ranker to generate $\bar{y}$
  - 33% noisy clicks on irrelevant docs

Presented $\bar{y}$	Q
A	q_1
B	q_2
Click	q_3
D	q_4
E	q_5
F	q_6
Click	q_7

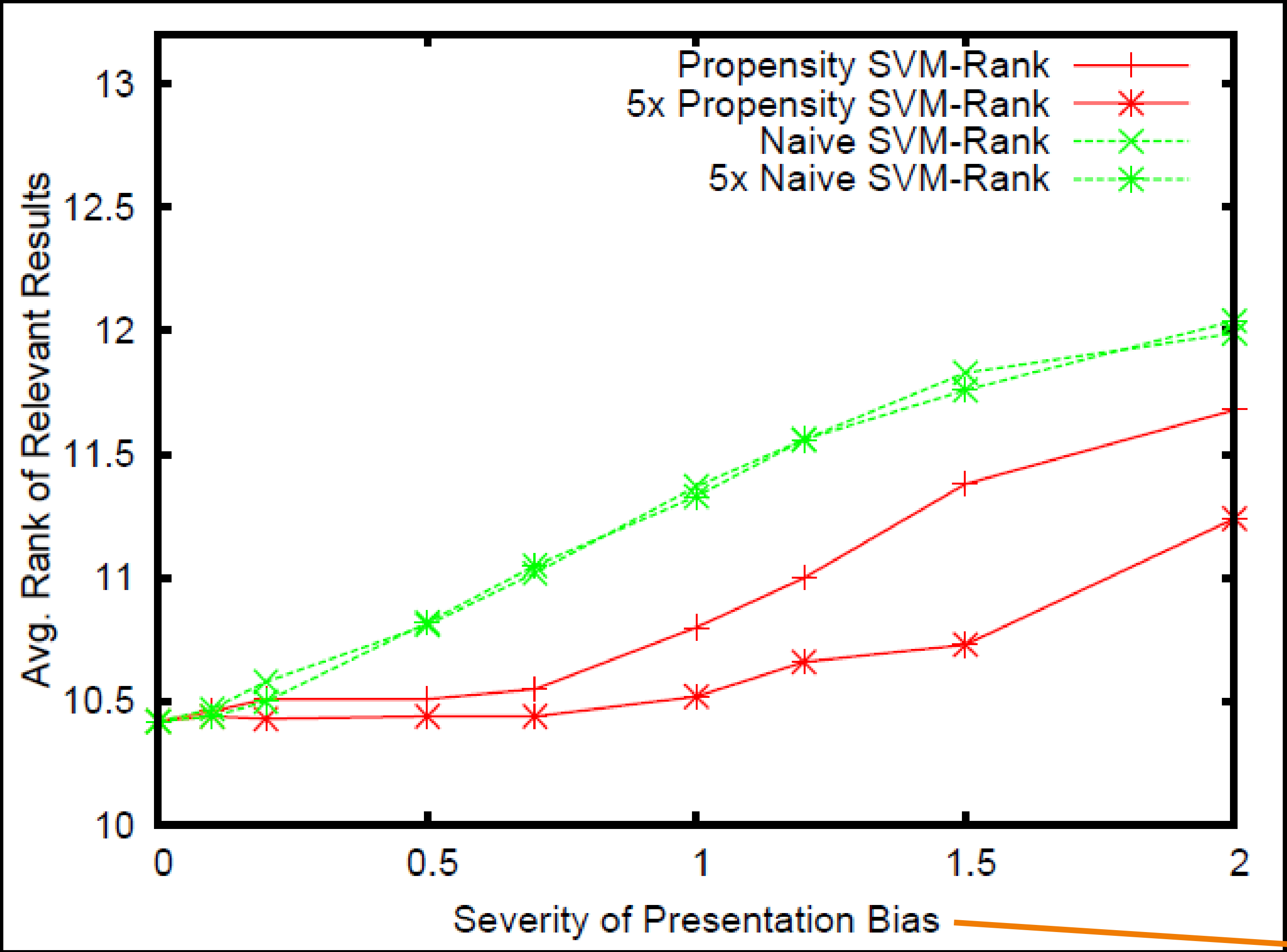
Scaling with Training Set Size



Clipping

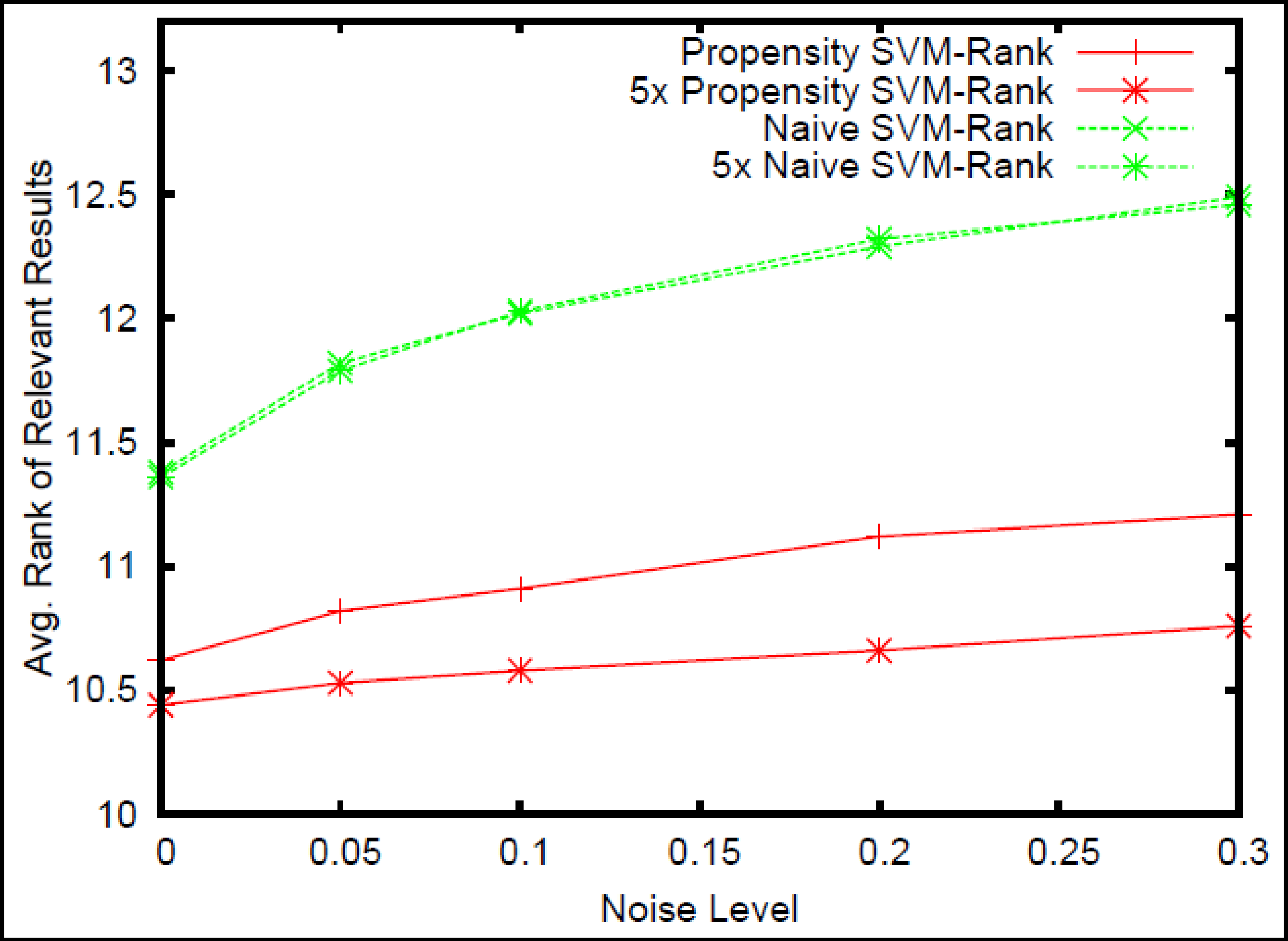

Production Ranker
Propensity SVM-Rank
Clipped Propensity SVM-Rank
Naive SVM-Rank
Noise-free Full-info Skyline
Avg. Rank of Relevant Results
13
12.5
12
11.5
11
10.5
10
1.7E3
1.7E4
1.7E5
Number of Training Clicks

Severity of Presentation Bias

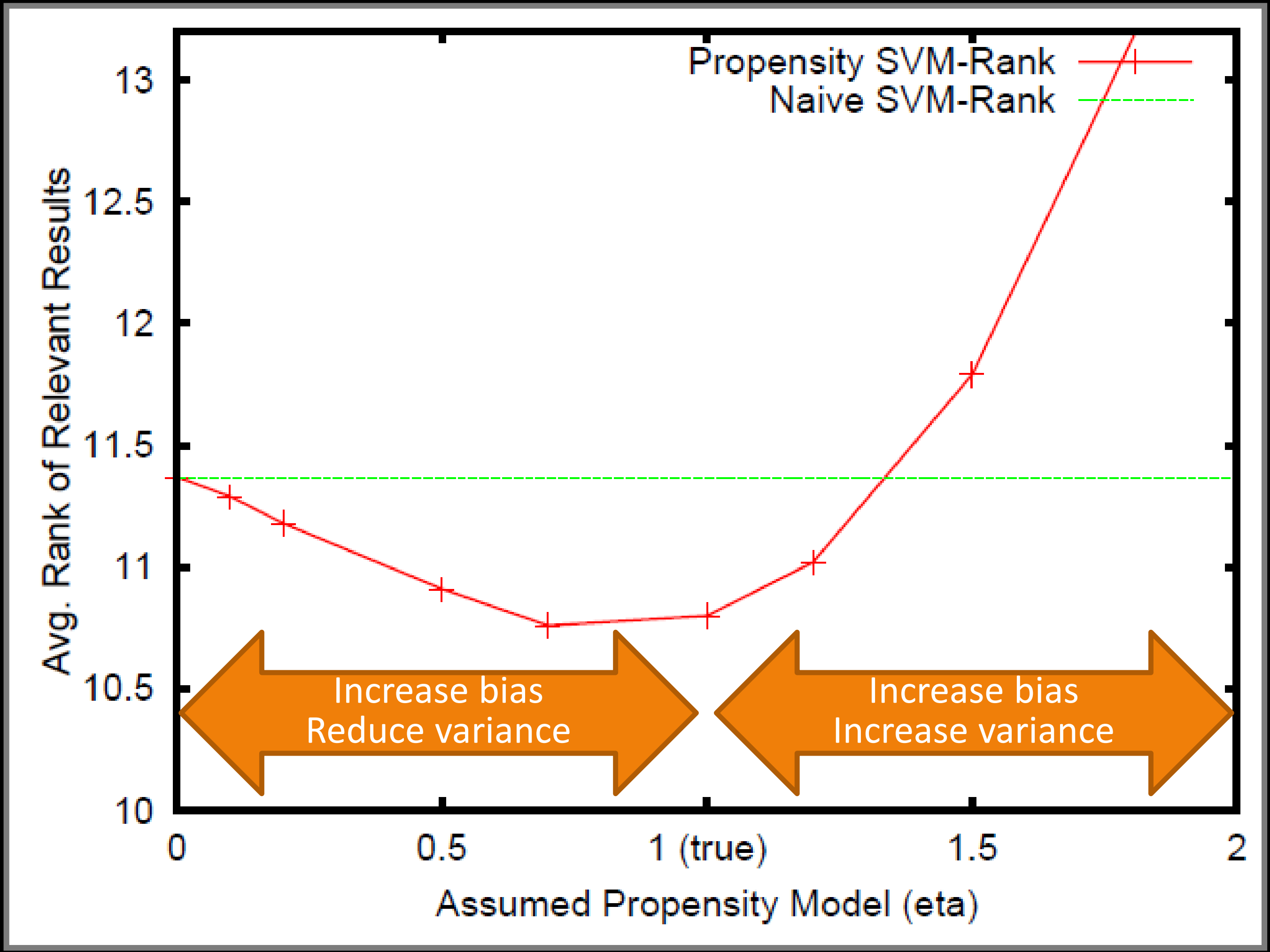


$$q_r = \left(\frac{1}{r}\right)^{\eta}$$

Increasing Click Noise



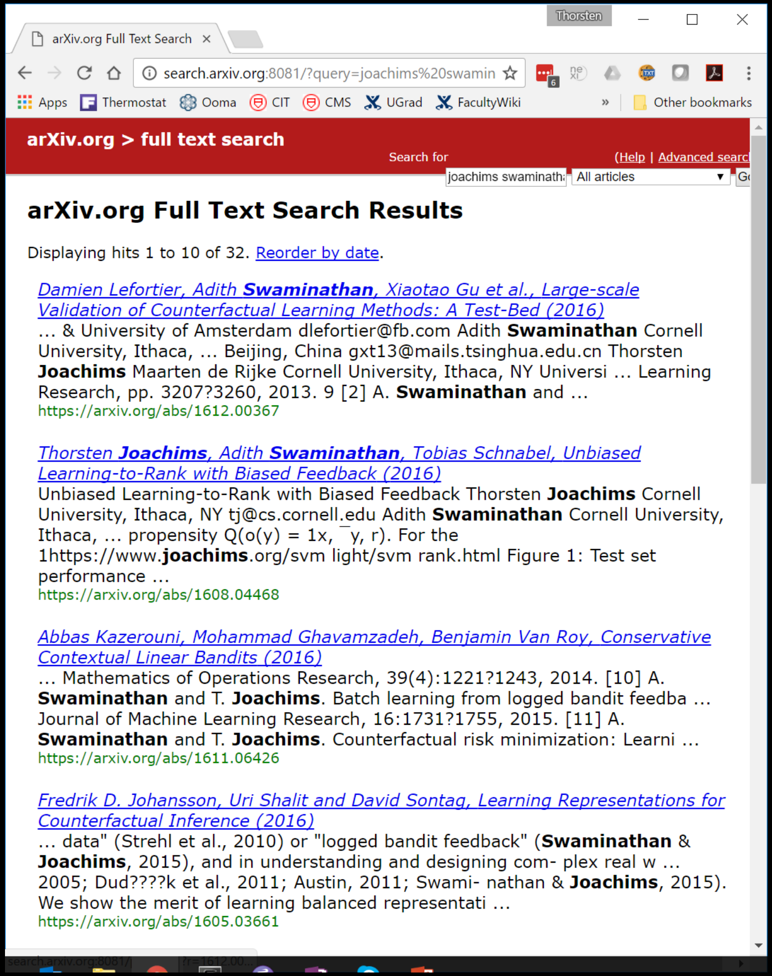
Misspecified Propensities



$$q_r = \left(\frac{1}{r}\right)^{\eta}$$

Real-World Experiment

- Arxiv Full-Text Search
  - Run intervention experiment to estimate q_r
  - Collect training clicks using production ranker
  - Train naïve / propensity SVM-Rank (1000 features)
  - A/B tests via interleaving



Interleaving Experiment	Propensity SVM-Rank wins	loses	ties
against Prod	87	48	83
against Naive SVM-Rank	95	60	102

Conclusions

- Partial-Information Learning to Rank
  - Selection bias is both interventional (π_0) and observational (user)
  - Combinatorial actions
- Approach
  - Decompose loss function into components
  - Get partial information about multiple losses
  - Unbiased estimate of each decomposed loss → ERM
- Open Questions
  - Propensity estimation beyond PBM and disruptive interventions
  - Other learning algorithms beyond Ranking SVM
  - Other counterfactual estimators beyond clipped IPS