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



Manually Labeled

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:

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

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

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

 \rightarrow Understand observation mechanism



Inverse Propensity Score Estimator

• Observation Propensities $Q(o_i = 1 | x, \overline{y}, r)$

- Random variable $o_i \in \{0,1\}$ indicates whether relevance label r_i for is observed
- Inverse Propensity Score (IPS) Estimator:

$$\widehat{\Delta}(y|r,o) = \sum_{i:o_i=1} \frac{rank(i|y) \cdot r_i}{Q(o_i = 1|\overline{y}, r)}$$
New Ranking
Need to know the propensities only for relevant/clicked docs.
$$= \sum_{i:o_i=1\wedge r_i=1} \frac{rank(i|y)}{Q(o_i = 1|\overline{y}, r)}$$

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

$$= \Delta(y|r)$$

$$= \Delta(y|r)$$

Presented \overline{y}

ERM for Partial-Information LTR

• Unbiased Empirical Risk:

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

Consistent Estimator of True Performance

• ERM Learning:

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

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 $Q(o_i = 1 | \overline{y}, r)$? \rightarrow Next week by Aman

BLBF vs. LTR

Batch Learning from Bandit Feedback

- Atomic actions
- Action y chosen by π_0 influences feedback
- Observe loss $\delta(x, y)$ for action y chosen by π_0 .
- Interventional → Logged propensities

Learning to Rank from Implicit Feedback

- Combinatorial actions
- Action y chosen by π_0 influences feedback
- Observe partial information about loss $\delta(x, y)$ for multiple y
- Interventional + Observational (user)

Propensity-Weighted SVM Rank

- Data: $S = (x_{j}, d_{j}, D_{j}, q_{j})^{n}$ Query Clicked Others Propensity $V_{\text{Query}} = \operatorname{argmin}_{w,\xi \ge 0} \frac{1}{2} w \cdot w + \frac{C}{n} \sum_{j} \frac{1}{q_{j}} \sum_{i} \xi_{j}^{i}$ $\forall \overline{d}^{i} \in D_{1} : w \cdot [\phi(x_{1}, d_{1}) - \phi(x_{1}, \overline{d}^{i})] \ge 1 - \xi_{1}^{i}$ \vdots $\forall \overline{d}^{i} \in D_{n} : w \cdot [\phi(x_{n}, d_{n}) - \phi(x_{n}, \overline{d}^{i})] \ge 1 - \xi_{n}^{i}$
- Loss Bound:

$$\forall w: rank(d, sort(w \cdot \phi(x, d)) \le \sum_{i} \xi^{i} + 1$$

[Joachims et al., 2002]

Position-Based Propensity Model

• Model:

$$P(c_i = 1 | r_i, rank(i | \overline{y})) =$$

$$P(o_i = 1 | rank(i | \overline{y}))$$

$$\cdot P(c_i = 1 | r_i, o_i = 1)$$

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

- Assumptions
 - Examination only depends on rank $\rightarrow Q(o_i = 1 | rank(i | \overline{y})) = q_r$
 - 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

svm - Google Search	× 🕀		
← ⇒ C fi ©	www.google.com/search?aq=f&gcx=c&sourceid=chrome&ie=UTI 😭 🤱 🍕	r,	
+You Web Images	Videos Maps News Shopping Gmail More - Sign in 🔅	Ĥ	
Google	svm		
Search	About 16,600,000 results (0.11 seconds)		
Everything Images Maps Videos	Support vector machine - Wikipedia. the free encyclopedia en wikipedia org/wiki/Support vector_machine A support vector machine (SWI) is a concept in statistics and computer science for is set of related supervised learning methods that analyze data and recognize Formal definition - History - Motivation - Linear SVM SVMI Summary for Silvercore Metals (or Ordinary - Yahoo) Einance	5	
News Shopping More	Innance yahoo com/q?s=SVM View the basic SVM stock chart on Yahool Finance. Change the date range, chart type and compare Silvercom Metals Inc Ordinary against other companies. SVM LP		
Any time Past hour Past 24 hours Past 2 days Past week Past month Past year Custom range	www.swmcards.net/ SVM. A leader in the grift card industry and devoted to helping your business reward, promote, entice and grow. Established in 1997, we handle the sales, SVM ANSEL Management - Home www.swmoline.co.uk/ Founded in 1990, SVM Asset Management is a privately-owned firm based in Edinburgh The three founding directors continue to own 100% of the equity, with UIRSVM – A Library for Support Vector Machines	'n	
All results Related searches More search tools	Allocation of the service of th		

Estimating the Propensities

• Experiment:

- Click rate at rank 1:
 - $q_1 \cdot E(r_i = 1 | rank(i | \overline{y}) = 1)$

Intervention:

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

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



 $\frac{1}{k}$ Click rate at rank k after swap

Click rate at rank 1

Experiments

Presented \overline{y}	Q
А	q_1
В	q_2
Click	q_3
D	q_4
Е	q_5
F	q_6
Click	q_7

- 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 \overline{y}
 - 33% noisy clicks on irrelevant docs

Scaling with Training Set Size



Clipping



Severity of Presentation Bias



Increasing Click Noise



Misspecified Propensities

 η



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

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



arXiv.org Full Text Search Results

Displaying hits 1 to 10 of 32. Reorder by date.

Damien Lefortier, Adith **Swaminathan**, Xiaotao Gu et al., Large-scale Validation of Counterfactual Learning Methods: A Test-Bed (2016)

... & University of Amsterdam dlefortier@fb.com Adith **Swaminathan** Cornell University, Ithaca, ... Beijing, China gxt13@mails.tsinghua.edu.cn Thorsten **Joachims** Maarten de Rijke Cornell University, Ithaca, NY Universi ... Learning Research, pp. 3207;3260, 2013. 9 [2] A. **Swaminathan** and ... https://arxiv.org/abs/1612.00367

Thorsten **Joachims**, Adith **Swaminathan**, Tobias Schnabel, Unbiased Learning-to-Rank with Biased Feedback (2016)

Unbiased Learning-to-Rank with Biased Feedback Thorsten **Joachims** Cornell University, Ithaca, NY tj@cs.cornell.edu Adith **Swaminathan** Cornell University, Ithaca, ... propensity Q(o(y) = 1x, -y, r). For the 1https://www.joachims.org/svm light/svm rank.html Figure 1: Test set performance ... https://arxiv.org/abs/1608.04468

Abbas Kazerouni, Mohammad Ghavamzadeh, Benjamin Van Roy, Conservative Contextual Linear Bandits (2016)

... Mathematics of Operations Research, 39(4):1221?1243, 2014. [10] A. **Swaminathan** and T. **Joachims**. Batch learning from logged bandit feedba ... Journal of Machine Learning Research, 16:1731?1755, 2015. [11] A. **Swaminathan** and T. **Joachims**. Counterfactual risk minimization: Learni ... https://arxiv.org/abs/1611.06426

Fredrik D. Johansson, Uri Shalit and David Sontag, Learning Representations for Counterfactual Inference (2016)

... data" (Strehl et al., 2010) or "logged bandit feedback" (Swaminathan & Joachims, 2015), and in understanding and designing com- plex real w ... 2005; Dud????k et al., 2011; Austin, 2011; Swami- nathan & Joachims, 2015). We show the merit of learning balanced representati ... https://arxiv.org/abs/1605.03661

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 \rightarrow ERM
- Open Questions
 - Propensity estimation beyond PBM and disruptive interventions
 - Other learning algorithms beyond Ranking SVM
 - Other counterfactual estimators beyond clipped IPS