Unbiased Learning-to-Rank with Biased Feedback

Thorsten Joachims, Adith Swaminathan, Tobias Schnabel International Conference on Web Search and Data Mining, 2017

> Presenter: Thorsten Joachims Dept of Computer Science, Cornell University



Interaction Logs: Search Engine

- Context *x*:
 - Query
- Action y:
 Ranking
- Reward/Loss $\Delta(y|x)$:
 - Search cost
 - Information gained
- Feedback:
 - Clicks on SERP

		S
svm - Google Search		
← ⇒ C ⋒ ©	www.google.com/search?aq=f&gcx=c&sourceid=chrome&ie=UTI 😭 🔝 🤜	6
+You Web Images	Videos Maps News Shopping Gmail More - Sign in 🔅	Â
Google	svm	
Search	About 16,600,000 results (0.11 seconds)	
Everything Images Maps Videos News	Support vector machine - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Support_vector_machine A support vector machine (SVM) is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize Formal definition - History - Motivation - Linear SVM SVM: Summary for Silvercorp Metals Inc Ordinary - Yahool Finance finance.yahoo.com/q?s=SVM	
Shopping More	view the basic SVM stock chart on Yanoo! Finance. Change the date range, chart type and compare Silvercorp Metals Inc Ordinary against other companies.	
<mark>Any time</mark> Past hour Past 24 hours	 SVM. A leader in the gift card industry and devoted to helping your business reward, promote, entice and grow. Established in 1997, we handle the sales, 	
Past 2 days Past week Past month Past year Custom range	SVM Asset Management - Home www.svmonline.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	
All results Related searches More search tools	LIBSVM A Library for Support Vector Machines www.csie.ntu.edu.tw/~cjlin/libsvm/ 5 Nov 2011 - An integrated and easy-to-use tool for support vector classification and regression.	
	· · · · · · · · · · · · · · · · · · ·	

Interaction Logs: Online Retail

- Context *x*:
 - Category
- Action *y*:
 Tile Layout
- Reward/Loss $\Delta(y|x)$:
 - Search cost
 - Product utility
- Feedback:
 - Purchases

E Etsy - Your place to buy at ×					
> C fi 🗋 www.ets	y.com		☆ 🕄		
Currency Converter 📵 Netta or	the net 🔣 DUS Wiki 💽 Thermos	tat 🔣 UGrad 🛅 AnxivSearch 🤅	🖲 CMS 🛛 » 🦲 Other bookmar		
ell Registry Community Blogs	Mobile Gift Cards				
Etsy Register	Sign In Search for items	s and shops	l		
Browse	TAT				
Art	What do brands and bloggers love on Ete				
Home & Living		Meet	Pages		
Jewelry	in the				
Women					
Men	Handpicked Items See mo	ore			
Kids					
Vintage	Anty Polline Cit	Series and the series of the s			
Weddings	11				
Craft Supplies	0.0	Distriction of the second seco	9.0		
Trending Items	12" Black, Silver, Gold & Oran	Fabric card holder - Party of	Cat, Black Cat photography,		
Halloween	gnomeswhi \$7.00 usp	octopurse \$13.50 usp	RikkiVanCamp \$25.00 usb		
Gift Ideas		~	1 1 1 1 1		
Mobile Accessories	AFRAIDA		~ ^		
Daily Finds Email	ethe DARK?		boo		
Get gift ideas, editors' picks & fresh trends in your inbox.	Halloween Card // Hand Lett EmDashPa \$4.50 uso	Halloween Art Spider Web W.	Halloween Boo Tote handst		
Enter your email					
Sign Up	494	1			
See our other newsletters.		S.J.	1 A		
More Ways to Shop					
Categories	The Prefect Dress	24 Halloween Ghost Fondan	10 Table Numbers, Cat Silh		
Gill Gards	Remarks \$57.76 Uen	LongeCaked \$16.95 List	MainaTita \$40.00 (ien		

Interaction Logs: Streaming Media

- Context *x*:
 - User
- Action *y*:
 Carousel layout
- Reward/Loss $\Delta(y|x)$:
 - Search cost
 - Enjoyment
- Feedback:
 - Plays



Learning-to-Rank from Clicks



Eye-Tracking

Eye tracking device





"Scanpath" output depicts pattern of movement throughout screen. Black markers represent fixations.

Detect and record where and what people look at

- Fixations: ~200-300ms;
 information is acquired
- Saccades: extremely rapid movements between fixations
- Pupil dilation: size of pupil indicates interest, arousal

How Many Links do Users View?



In Which Order are Results Viewed?



=> Users tend to read the results in order

Examination Curve from Eyetracking



Position \rightarrow Exposure \rightarrow Feedback

[Granka et al., 2007]

Outline

- Learning-to-Rank from User Interactions - Find new ranking policy π that selects y with better δ
- Batch Learning-to-Rank from Partial Labels
 - Learning from partial and biased feedback
 - Learning Principle: Unbiased Partial-Information ERM
 - Learning Algorithm: Propensity SVM-Rank
 - Propensity Estimation for Ranking
 - Break confounding through position randomization
 - Intervention Harvesting
 - Contextual propensity models

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:

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

- 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:

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

New Ranking

• Unbiasedness: $E_o[\widehat{\Delta}(y \mid r, o)] = \Delta(y \mid r)$

Presented \overline{y}	Q
А	1.0
В	0.8
С	0.5
D	0.2
Е	0.2
F	0.2
G	0.1

ERM for Partial-Information LTR

Unbiased Empirical Risk:

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

Consistent Estimator of True Performance

• ERM Learning:

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

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

Propensity-Weighted SVM Rank

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

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

[Herbrich at al., 1999] [Joachims et al., 2002] [Joachims et al., 2017]

Position-Based Propensity Model

• Model:

$$P(c_{i} = 1 | r_{i}, rank(i | \overline{y})) = q_{rank(i | \overline{y})} \cdot [r_{i} = 1]$$

- Assumptions
 - Examination only depends on rank
 - Click reveals relevance if rank is examined

Presented \overline{y}	Q
А	q_1
В	q_2
С	<i>q</i> ₃
D	q_4
E	q_5
F	q_6
G	q_7

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

Scaling with Training Set Size



Scaling with Training Set Size



Severity of Presentation Bias



 q_r

Increasing Click Noise



Misspecified Propensities



 q_r

Outline

- Learning-to-Rank from User Interactions — Find new ranking policy π that selects y with better δ
- Batch Learning-to-Rank from Partial Labels
 - Learning from partial and biased feedback
 - Learning Principle: Unbiased Partial-Information ERM
 - Learning Algorithm: Propensity SVM-Rank
- Propensity Estimation for Ranking
 - Control for relevance through position randomization

Position-Based Propensity Model

• Model:

$$P(c_{i} = 1 | r_{i}, rank(i | \overline{y})) = q_{rank(i | \overline{y})} \cdot [r_{i} = 1]$$

- Assumptions
 - Examination only depends on rank
 - Click reveals relevance if rank is examined

Presented \overline{y}	q
А	q_1
В	q_2
С	<i>q</i> ₃
D	q_4
E	q_5
F	q_6
G	q_7

Examination Curve from Eyetracking



Estimating the Propensities

Idea: Randomization to control for relevance
 → Swap Interventions



Real-World Experiment

- Arxiv Full-Text Search
 - Run Swap(1,r) 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 and Discussion

- Learning to Rank from User Interactions
- Batch Learning-to-Rank from Partial Labels
 - Find new ranker π that selects y with improved rank metric
 - Positive-only feedback on subset of items
 - Correct for biased feedback due to bias in user exposure
 - Estimate propensities by controling for relevance through swap interventions
- What is still missing?
 - Improve on simplistic propensity model
 - How to deal with zero propensities
 - Biases that do not work through exposure (e.g. Trust Bias)
 - Other learning algorithms and ranking metrics
 - Etc.