Counterfactual Evaluation and Learning from Logged User Feedback

Adith Swaminathan

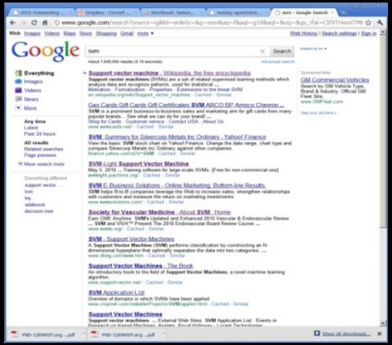
adith@cs.cornell.edu

Department of Computer Science

Cornell University

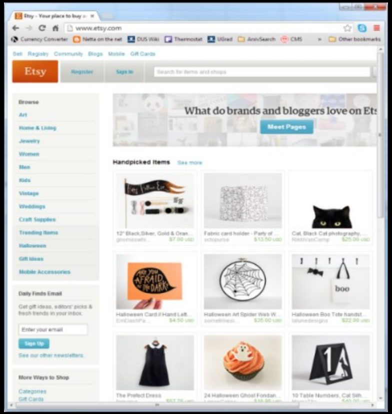
Committee: Thorsten Joachims (chair), Johannes Gehrke, Éva Tardos, Robert Kleinberg

Interactive Systems



Search

E-commerce



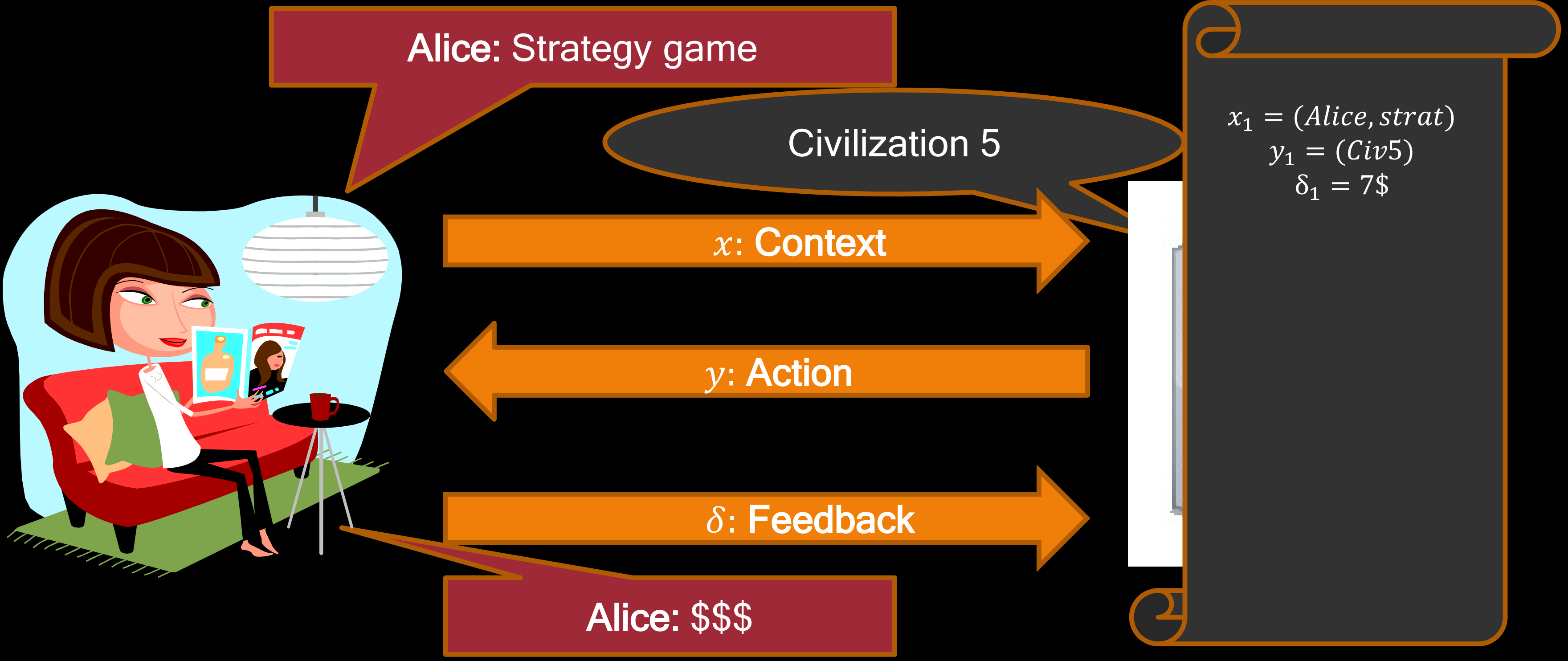
We collect user interactions for

- Personalization
- Evaluating performance
- Improving systems …

Entertainment

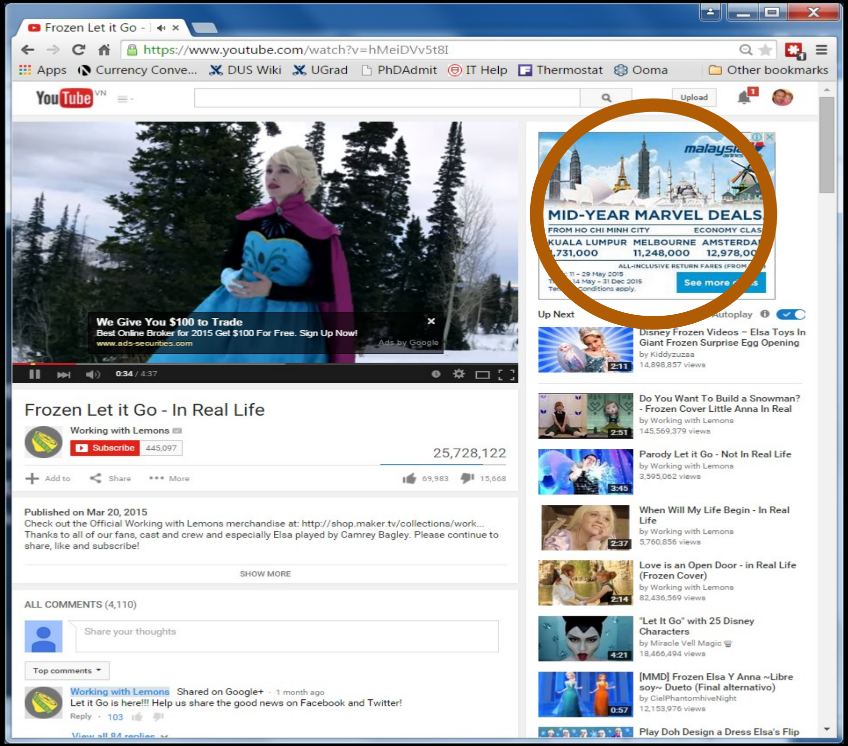
SmartHome …

Logs of Interactive Systems



Example: Ad Placement

- Context x:
  - User and page
- Action y:
  - Ad
- Feedback δ:
  - Click / no-click



Example: Chatbot

- Context x:
  - Query
- Action y:
  - GIF
- Feedback δ:
  - Thumbs up / down

http://peeqo.com/

Goal

Use

$$x_1 = (Alice, strat)$$
$$y_1 = (Civ5)$$
$$\delta_1 = 7\$$$

…

to

- Evaluate new model
- Train models

Non-solutions

- Ignore

$$x_1 = (Alice, strat)$$
$$y_1 = (Civ5)$$
$$\delta_1 = 7\$$$
…

- Evaluate: Online A/B
- Train: Bandit algorithms

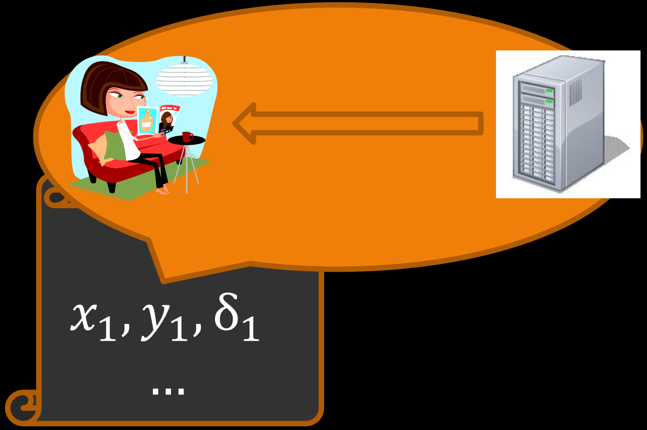
- Annotate

$$x_1 = (Alice, strat)$$
$$y_1 = (Civ5)$$
$$\delta_1 = 7\$$$
$$y^* = (AoE2)$$
…

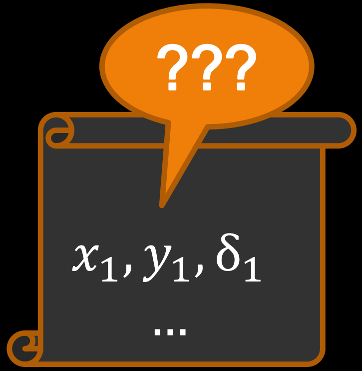
- Evaluate: Hold-out
- Train: Supervised learning

Thesis: Re-use logged data

- Logged interventions

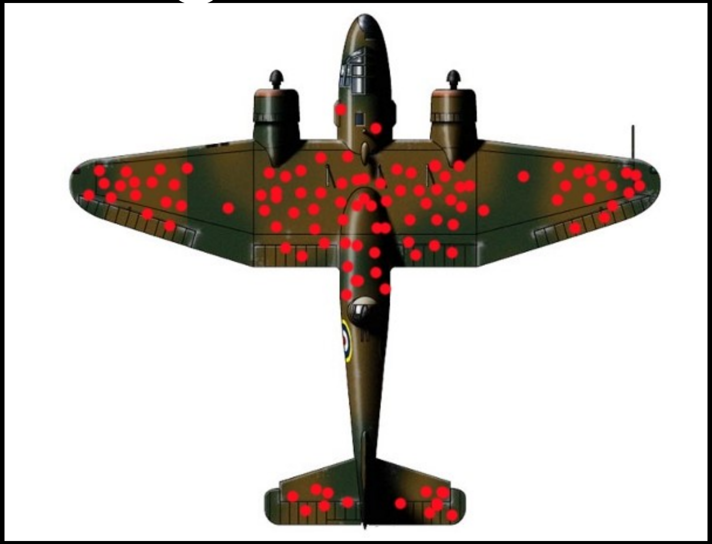


- Observational logs



- Evaluating slates [ICML'16 workshop]
- Batch Learning from Bandit Feedback (BLBF) [ICML'15]
- Better BLBF [NIPS'15]
- Recommender systems [ICML'16]
- Learning to rank [WSDM'17]

Wald's insight: What's missing?



- Where to add armor? Cover bullet-holes? (Survivor bias!)
- Beware: Confounding due to missing info

Overview

- Introduction
  - Interactive systems
  - Reusing data
- "How good is this new system?"
  - Project: MNAR (Schnabel et al, ICML 2016)
- "Find the best new system"
  - Project: POEM

Movie Recommendation

O
Observed Y/N

	Horror					Romance					Drama				
Horror Lovers	5		5			1							3		
		5	5												
		5		5		1					3				
	5		5		5								3		
Romance Lovers								5		5					
							5								
						5			5	5	3				
	1						5	5						3	

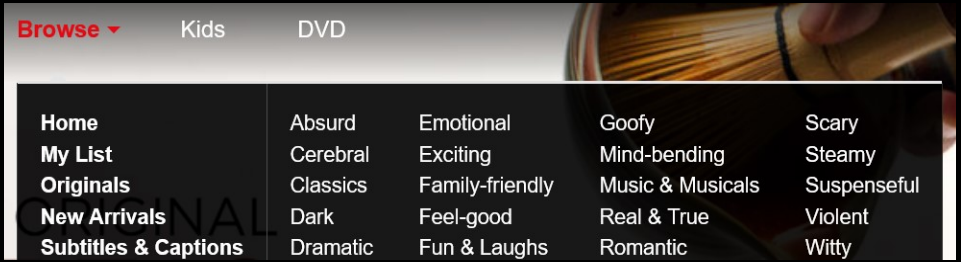
Data is Missing Not At Random (MNAR)

Y
True Rating

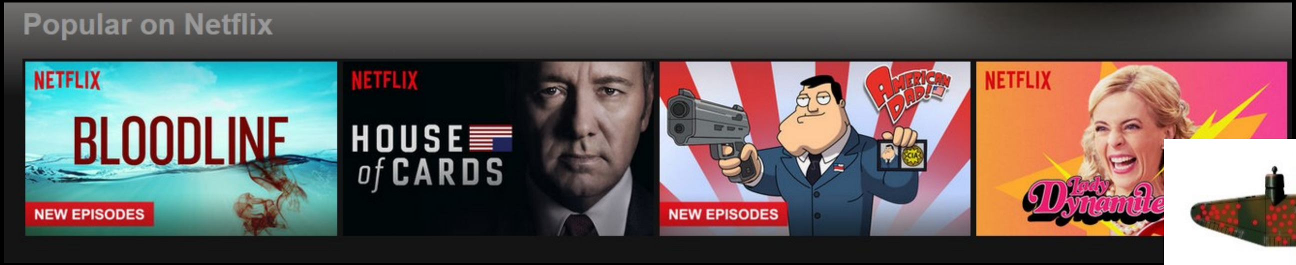
Example adapted from (Steck et al, 2010)

Selection Bias in Recommendations

- User-induced (e.g. browsing)

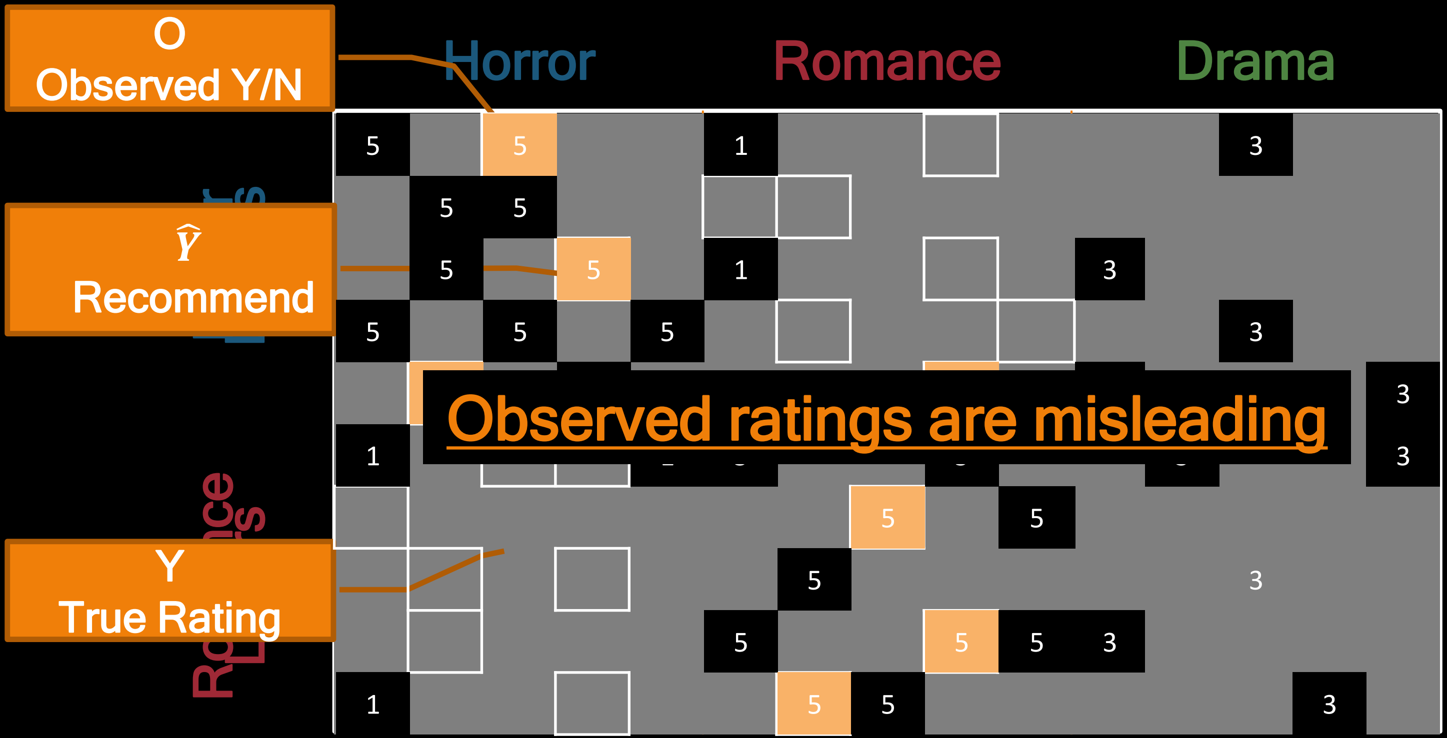


- System-induced (e.g. advertising)



Question: What if we ignore these biases?

Evaluating recommendations under Selection Bias



Evaluating rating predictions under Selection Bias

	Horror					Romance					Drama				
Horror Lovers	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Romance Lovers	5	5	5	5	5	5	5	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3

$\hat{Y}_1$
Pred Ratings (worse)

	Horror					Romance					Drama				
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	5	5	5	5	5
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5

$\hat{Y}_2$
Pred Ratings (better)

Observed losses are misleading

Recommendations as Treatments

Fix selection bias → potential outcomes framework

Counterfactual Outcomes Y

~~Items~~ treatments

~~Users~~ patients

5	1	3
1	5	3

Factual Outcomes $\tilde{Y}$

5		5			1							3		
	5	5												
	5		5		1					3				
5		5		5								3		
	5		5					1		3				3
1				1	5			5			3			3
							5		5					
						5								
					5			5	5	3				
1						5	5						3	

⇒ Understand **assignment mechanism**

(Imbens & Ruben, 2015)

Assignment Mechanism for Recommendation

$$P_{u,i} = P(O_{u,i} = 1)$$

Inverse Propensity Scoring (IPS) is unbiased if $P_{u,i} > 0$:

$$\hat{R}_{IPS} = \frac{1}{U \cdot I} \sum_{(u,i)} \frac{\mathbb{1}\{O_{ui}=1\}}{P_{u,i}} \left(Y_{u,i} - \hat{Y}_{u,i}\right)^2$$

Propensities P

Horror	Romance	Drama
p	$p/10$	$p/2$
$p/10$	p	$p/2$

(Horvitz & Thompson, 1952; Rosenbaum & Rubin, 1983; ...)

Interventional vs. Observational

- Interventional (Controlled Experiments)
  - We control assignment mechanism (e.g. ad placement)
  - Propensities $P_{u,i} = P\left(O_{u,i} = 1\right)$ known [Just log propensities!]
  - Requirement: $P_{u,i} > 0$ (prob. assignment)

- Observational
  - Assignment mechanism not under our control (e.g. reviews/ratings)
  - Use features Z; $\hat{P}_{u,i} = P\left(O_{u,i} = 1 | Z\right)$ [Estimate propensity]
  - Requirement: $O_{u,i} \perp Y_{u,i} \mid Z$ (unconfounded)

Propensity Estimation

- Supervised Regression Problem

$$\hat{P}_{u,i} = P\left(O_{u,i} = 1 \mid Z\right)$$

- Off-the-shelf ML, e.g.,
  - Logistic regression
  - Naïve Bayes
  - Bernoulli Matrix Factorization
  - …

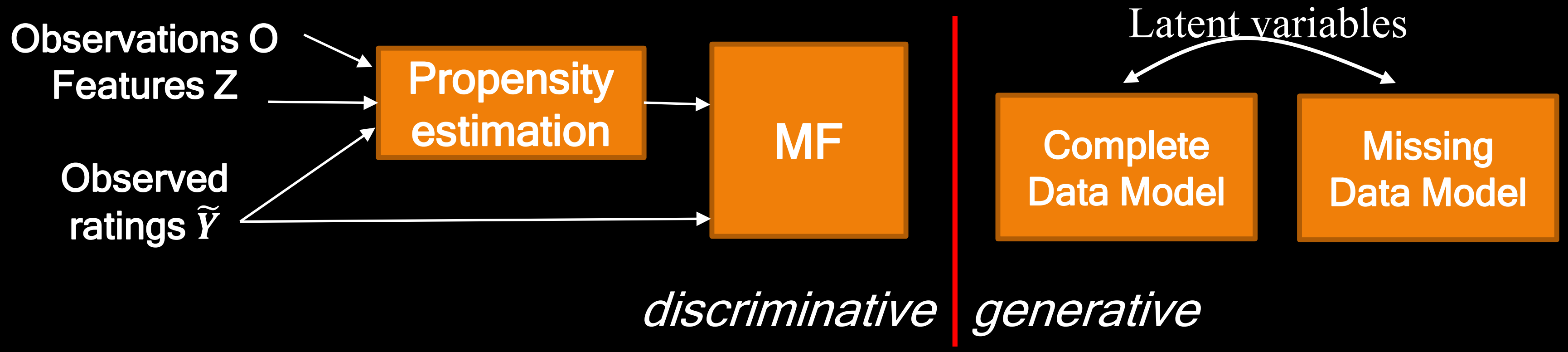
Observations O

Horror					Romance					Drama				
1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1
1	0	0	0	1	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	1	0

IPS is robust to inaccurate propensities

Debiased Collaborative Filtering

$$\hat{Y}^{ERM} = \operatorname*{argmin}_{V,W} \left\{ \sum_{O_{u,i}=1} \frac{1}{P_{u,i}} \left(Y_{u,i} - V_u W_i\right)^2 + \lambda\left(\|V\|_F^2 + \|W\|_F^2\right) \right\}$$



(Marlin et al, 2007; Steck, 2011; ...)

Collaborative Filtering Results

- Two real-world MNAR datasets
  - YAHOO: Song ratings (15400 users; Marlin & Zemel, 2009)
  - COAT: Shopping ratings (300 users; new Schnabel et al, 2016)
- Report performance on MAR datasets

	YAHOO		COAT	
	MAE	MSE	MAE	MSE
MF-IPS	**0.810**	**0.989**	**0.860**	**1.093**
MF-Naive	1.154	1.891	0.920	1.202
HL MNAR	1.177	2.175	0.884	1.214
HL MAR	1.179	2.166	0.892	1.220

http://www.cs.cornell.edu/~schnabts/mnar/

Overview

- Introduction
  - Interactive systems
  - Reusing data
- "How good is this new system?"
  - Project: MNAR
- "Find the best new system"
  - Project: POEM

(Swaminathan & Joachims, ICML 2015)

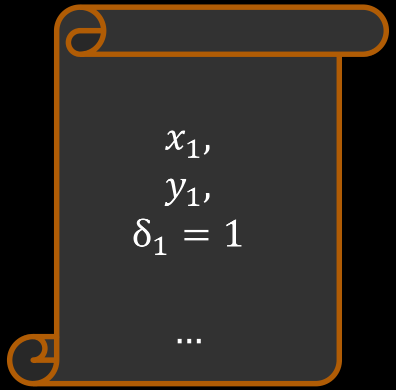
Evaluate with logged interventions

- New System π,
- How good is π?

$$\pi: x \mapsto y$$

$$\mathcal{R}(\pi) = \mathbb{E}_x \mathbb{E}_\pi[\delta]$$

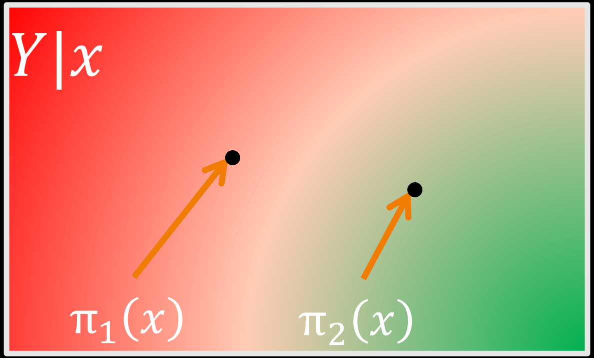
- Estimate $\mathcal{R}(\pi)$ using



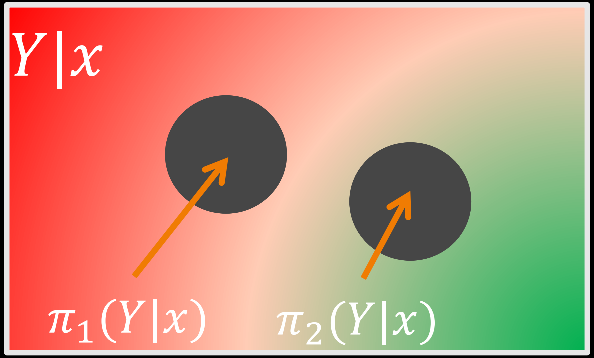
- $y_i \sim$ Old System π_0

System: Stochastic Policies

- Definition [*Deterministic Policy*]:
  Function
  $$y = \pi(x)$$
  that picks action y for context x

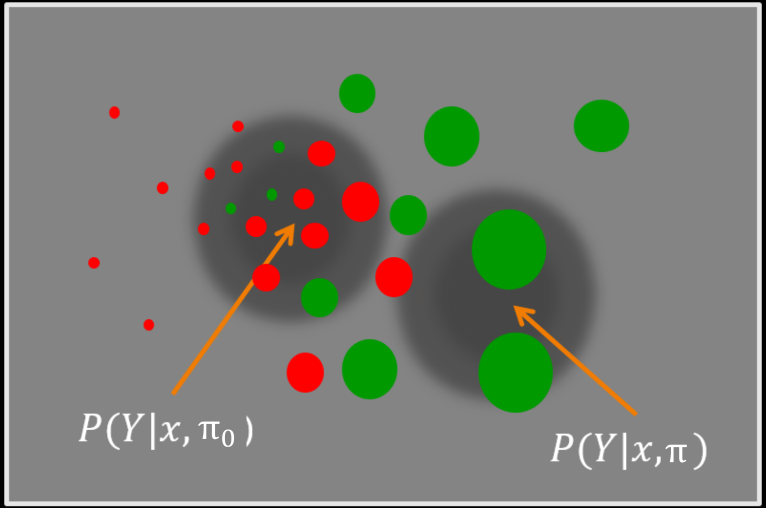


- Definition [*Stochastic Policy*]:
  Distribution
  $$\pi(y|x)$$
  that samples action y given context x



Recap: IPS

$$\hat{\mathcal{R}}_{ips}(\pi) = \frac{1}{n}\sum_{i}\frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)}\delta_i$$



Now: Learning Problem

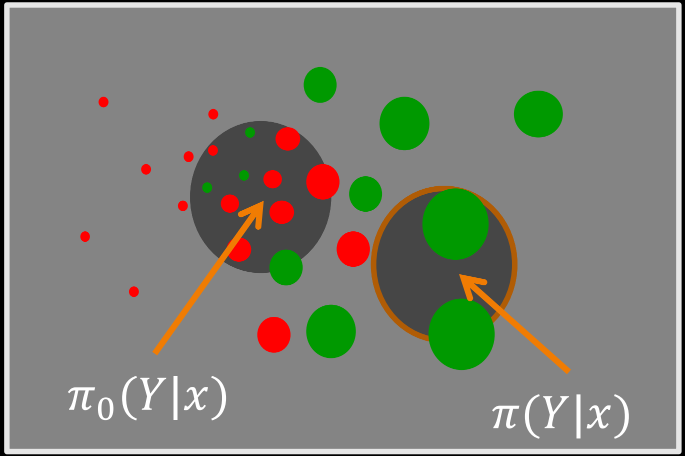
- Find the "best" $\pi^* \in \Pi$, $\quad \underset{\pi \in \Pi}{\operatorname{argmin}} \mathcal{R}(\pi)$

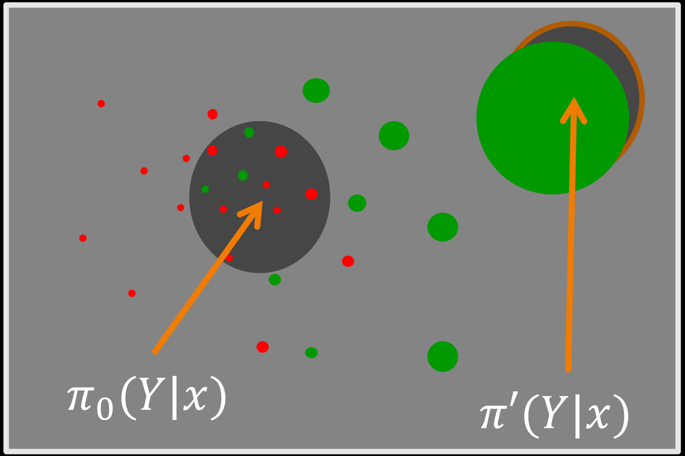
- Empirical Risk Minimization: Work-horse of ML

$$\pi^* = \underset{\pi \in \Pi}{\operatorname{argmin}} \;\; \hat{\mathcal{R}}(\pi) + \lambda \; \mathrm{Reg}(\pi)$$

Train loss Regularizer

ERM with IPS: Issue





$$\underset{\pi}{\operatorname{argmin}}\, \hat{\mathcal{R}}(\pi) = \frac{1}{n}\sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)}\delta_i$$

Can we detect and avoid IPS failure when learning?

ERM: Generalization Error Bound

Classic ERM: $\underset{\pi \in \Pi}{\text{argmin}} \quad \hat{\mathcal{R}}(\pi) \ + \quad \lambda \, \text{Reg}(\pi)$

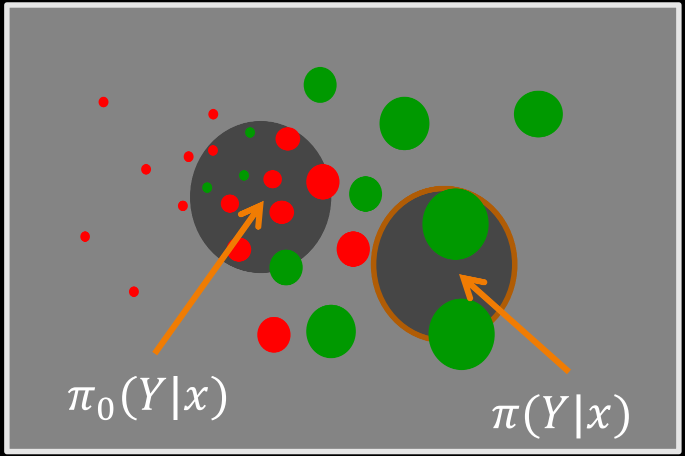
Train loss Regularizer

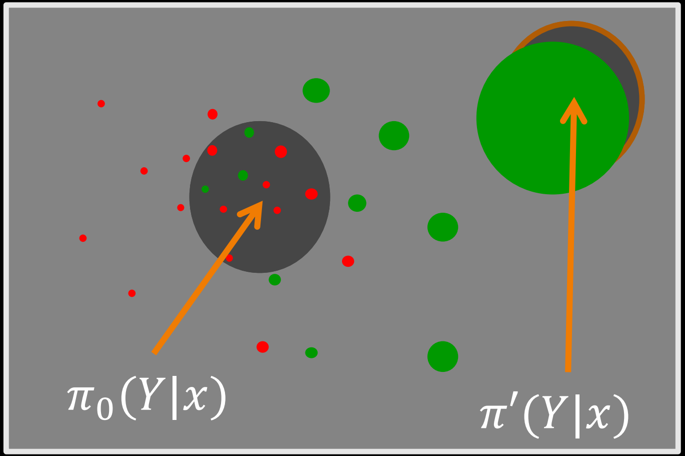
Classic Risk Bound: $\mathcal{R}(\pi) \leq \quad \hat{\mathcal{R}}(\pi) + \quad O(\mathrm{C}[\Pi])$

Data used to estimate $\hat{\mathcal{R}}(\pi)$ did not depend on π

(Vapnik & Chervonenkis, 1979)

Now: π influences its data





$$\underset{\pi}{\operatorname{argmin}} \hat{\mathcal{R}}(\pi) = \frac{1}{n} \sum_i \frac{\pi(y_i|x_i)}{\pi_0(y_i|x_i)} \delta_i$$

Counterfactual Learning

Risk Bound: $\mathcal{R}(\pi) \leq \quad \hat{\mathcal{R}}(\pi) \; + \; O\left(\sqrt{\frac{\widehat{Var}(\pi)}{n}}\right) + \; O(\mathrm{C}[\Pi])$

Off-policy est.   Emp. variance   Regularizer

Objective: $\underset{\pi \in \Pi}{\operatorname{argmin}} \quad \hat{\mathcal{R}}(\pi) \; + \; \lambda_1 \sqrt{\frac{\widehat{Var}(\pi)}{n}} \; + \lambda_2 \; \mathrm{Reg}(\pi)$

Counterfactual Risk Minimization

Accounts for different $\pi(y|x)/\pi_0(y|x)$ variability across Π

(Maurer & Pontil, 2009; Bottou et al, 2013; Swaminathan & Joachims, 2015)

CRM for Structured Prediction

Policy class, H: Stochastic linear rules

$$\pi_w(y|x) \propto \exp\{w^T \psi(x,y)\}$$

Same form as Cond. Random Field or Structural SVM

Learning: Use $\langle x_i, y_i, \delta_i, p_i \rangle$ to find good w

Policy Optimizer for Exponential Models (POEM)

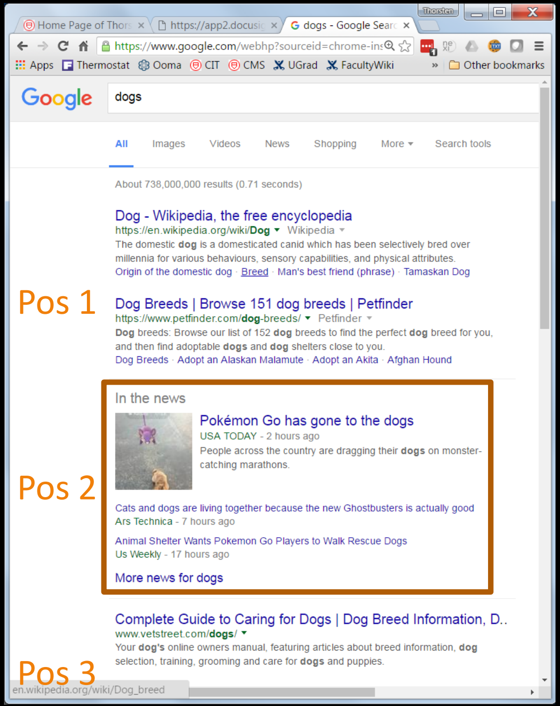
http://www.cs.cornell.edu/~adith/POEM/

(Swaminathan & Joachims, 2015)

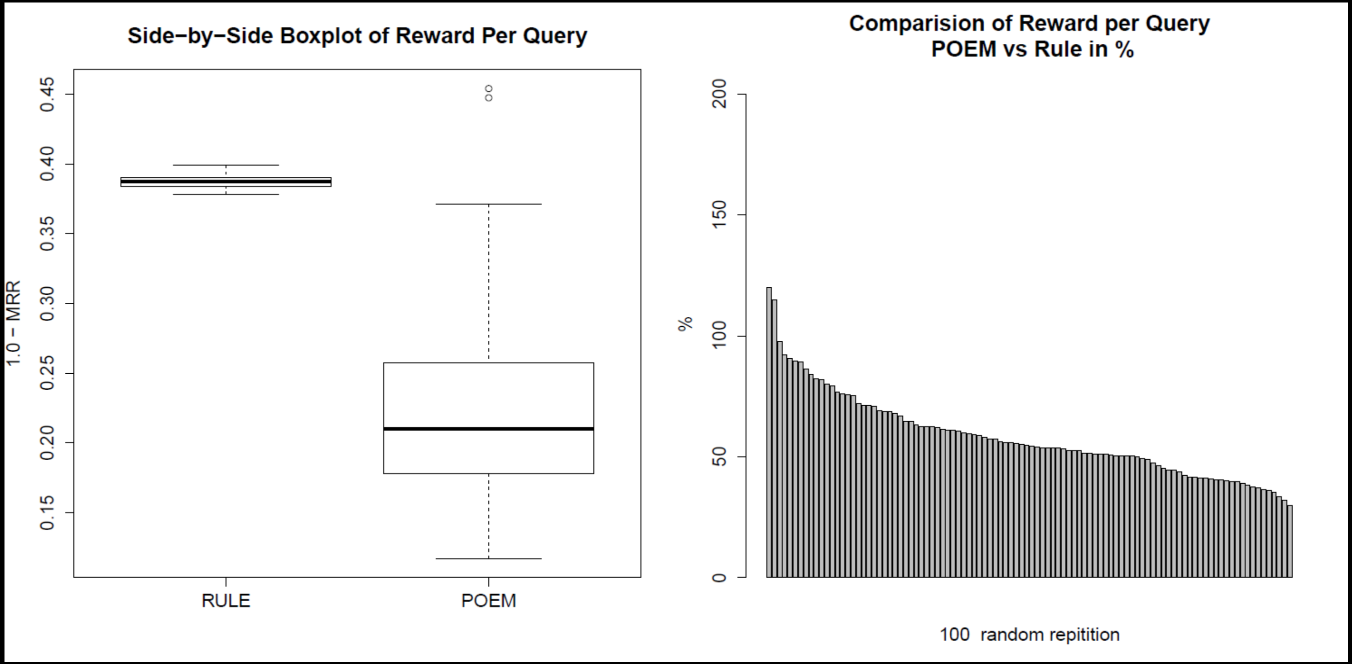
Case Study: Newsbox Placement

- Context x: Query, User, Ranked docs, Newsbox content features
- Action y: Position to place newsbox
- Loss δ: (1 - MRR) of entire SERP

- Logger π_0: Multinomial using production position scorer



News Box Placement: Results



Across 100 datasets, POEM consistently beats production ranker

Overview

- Introduction
  - Interactive systems
  - Reusing data
- “How good is this new system?”
- “Find the best new system”

- Discussion

Goal

Re-use

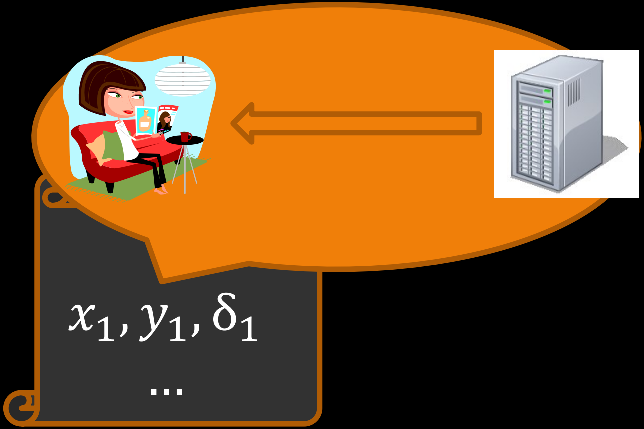
$$x_1 = (Alice, strat)$$
$$y_1 = (Civ5)$$
$$\delta_1 = 7\$$$
…

to

- Evaluate new model ***offline*** — IPS (Off-policy estimation)
- Train models ***offline*** — POEM

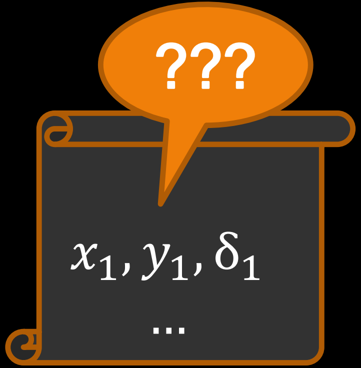
Thesis: Re-use logged data

- Logged interventions



- Evaluating slates [ICML'16 workshop]
- Batch Learning from Bandit Feedback (BLBF) [ICML'15]
- Better BLBF [NIPS'15]

- Observational logs



- Recommender systems [ICML'16]
- Learning to rank [WSDM'17]

Thesis: Other Projects

Better BLBF

- Equivariant Optimization with Self-Normalized Estimators
  - (Swaminathan & Joachims, NIPS 2015)
  - Solve propensity overfitting when learning from bandit feedback

Several variance reduction methods (clipping, control variates, explicit control, under-fit propensities). How do they compare and interact?

Thesis: Other Projects

Evaluating slates

- Off-Policy Evaluation for Slate Recommendation
  - (Swaminathan et al, ICML Workshop on Personalization 2016)
  - Usable alternative to IPS for combinatorial contextual bandits

Several ways to achieve a good bias-variance trade-off. How best to explore/exploit structured action spaces?

Thesis: Other Projects

Learning to rank

- De-biasing Learning to Rank when using Implicit Feedback
  - (Joachims et al, WSDM 2017)
  - To get around getting randomized data, piggyback on randomness in user actions

Can employ click models as propensity estimators. How to best adapt them for this purpose?

Counterfactual Techniques

"Model the world"

e.g., predict user-item ratings
Project: MNAR

Causal modelling

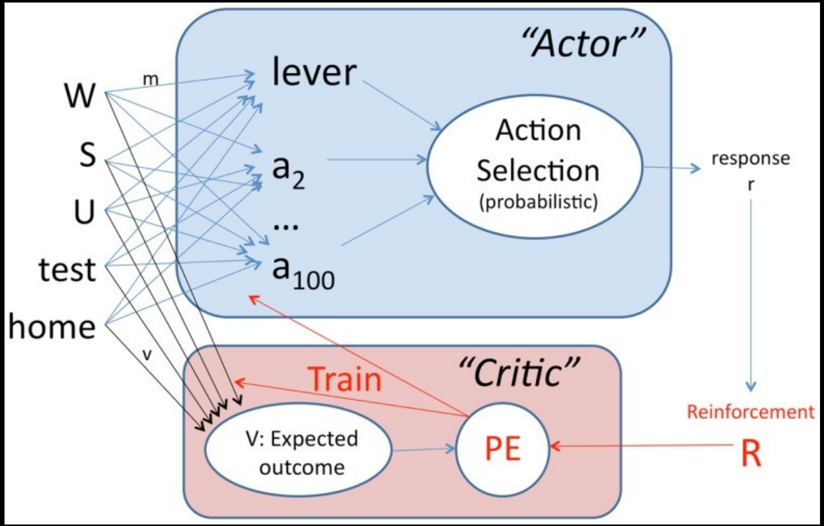
"Model the bias"

e.g., CTR of recommender
Project: POEM

Monte Carlo estimation

Looking Ahead

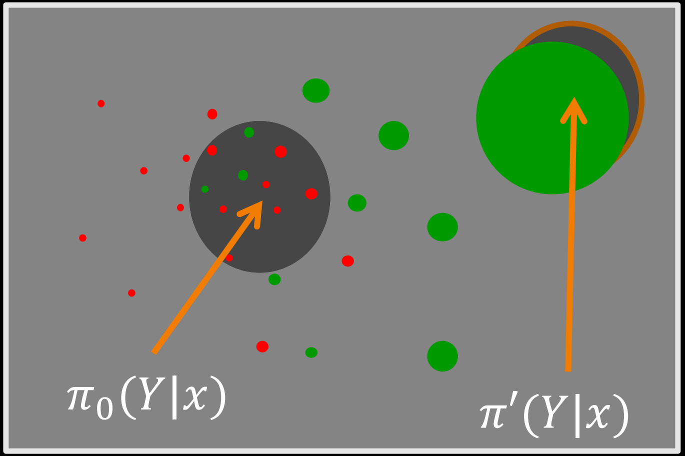
- Bridge Causal Inference and ML
- Off-policy Reinforcement Learning



Conclusion

See also http://www.cs.cornell.edu/~adith/Criteo/ and http://www.cs.cornell.edu/~adith/CfactSIGIR2016/index.html

Horror					Romance					Drama				
1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1
1	0	0	0	1	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	1	0



Thanks!

adith@cs.cornell.edu

Thanks!

- Thorsten
- Johannes
- Bobby
- Éva
  - Stavros
  - Telang
  - Karn
  - Z
  - Auro
  - Basu
  - Chenhao
  - Jon
  - Yexiang
  - Lu
- Ashwinkumar
- Wenlei
- Raghu
- Prathmesh
- Omkar
- Sunny
- Hooda
- Karthik Raman
- Siddarth Chandrasekaran
  - Saraswathi
  - Swaminathan
  - Arun
  - Abi
  - Indu
  - Prema
  - Raja
  - Hema
  - Ramesh
  - Ria/Nina
  - Lillian
  - Becky
  - Charlie
  - K.P.
  - M.
  - Abbi
  - Ashish
  - Harshal
  - Shinde
  - Binit
  - Reetika
  - Anne Schuth
  - Artem
  - Daan
- Hema Swetha
- Ashesh
- Abhishek
- Anshumali
- Shuo
- Amit
- Nikos
- Tobias
- Ashudeep
- Isaac
- Laure
- Matthew
- Fabian
- Ethan
- Soumen
- Pushpak
- Aditya GP
- Ramdas
  - Sankaranarayanan
  - Aravind
  - Moontae
  - Vikram
  - Aparna
  - Karthik Sridharan
  - Sarah
  - Sumita
  - Tirth
  - Maithra
  - Claire
  - David Mimno
  - Ainur
  - Bishan
  - Arzoo
- Ruben
- Pannaga
- Krishnaram
- Anitha
- David Grangier
- Navdeep
- Peter Frazier
- AkshayK/B
- Alekh
- Miro
- John
- Maarten
- Damien Lefortier