

# Learning from Rational\* Behavior

Josef Broder, Olivier Chapelle, Geri Gay, Arpita Ghosh,  
Laura Granka, Thorsten Joachims, Bobby Kleinberg, Madhu  
Kurup, Filip Radlinski, Karthik Raman, Tobias Schnabel,  
Pannaga Shivaswamy, Adith Swaminathan, Yisong Yue

Department of Computer Science  
Department of Information Science  
Cornell University

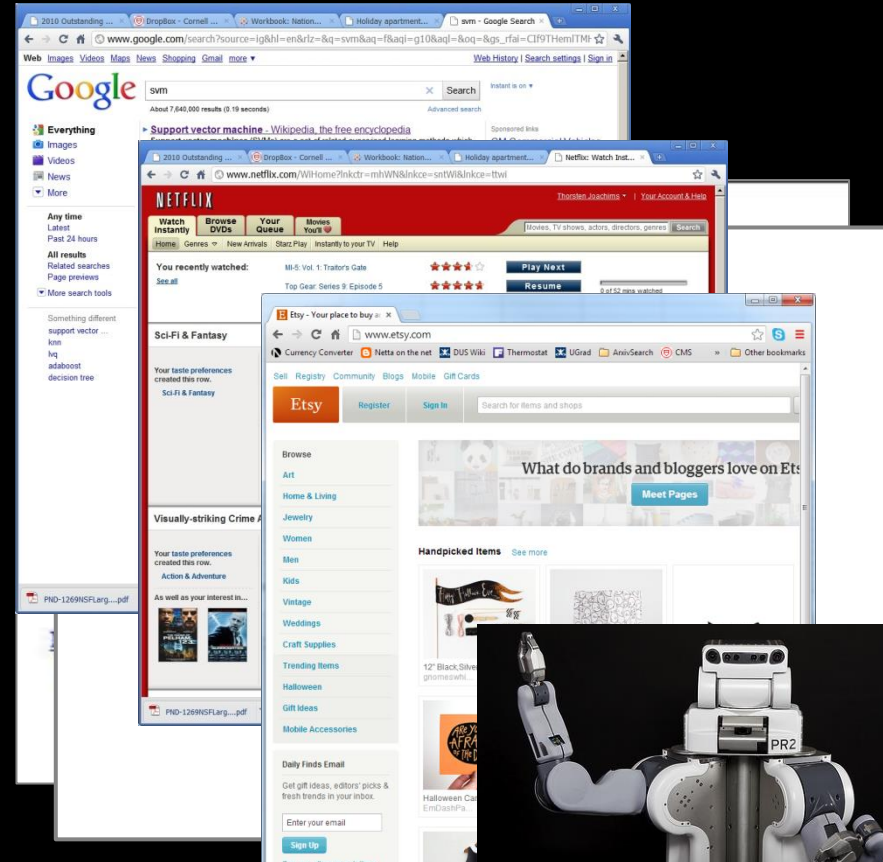
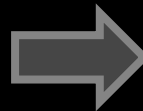
\*) Restrictions apply. Some modeling required.

# Knowledge-Based Systems

Text Understanding in Lilog:  
Integrating Computational  
Linguistics and Artificial Intelligence  
Final Report on the IBM Germany  
Lilog-Project (Lecture Notes in  
Computer Science)



Herzog, O.



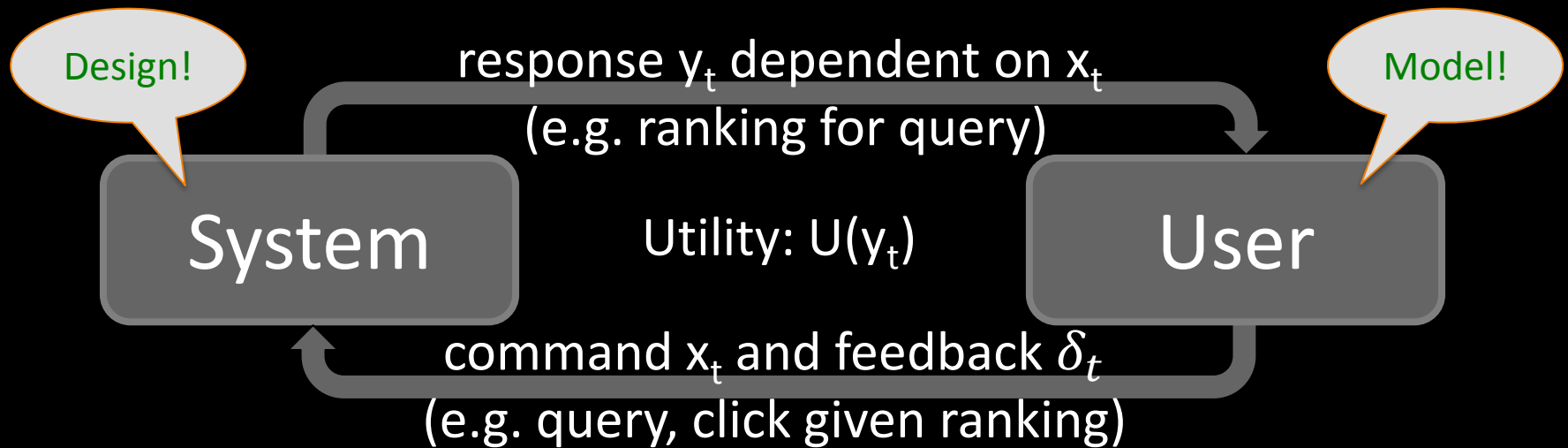
Users as a Source of Knowledge

# Supervised Batch Learning

- Data:  $(x, y^*) \sim P(X, Y^*)$ 
    - $x$  is input
    - $y^*$  is true label
  - Rules:  $f \in H$
  - Prediction:  $\hat{y} = f(x)$
  - Loss function:  $\Delta(\hat{y}, y^*)$
- Find  $f \in H$  that minimizes prediction error

$$R(f) = \int \Delta(f(x), y^*)$$

# Interactive Learning System



- Observed Data  $\neq$  Training Data in Naïve Learning Model
  - Observed data is user's decisions
  - Decisions need proper interpretation
- New Learning Models that Integrate User and Algorithm
  - Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm

# Decide between two Ranking Functions

Distribution  $P(x)$   
of  $x=(\text{user, query})$

$\vdots$   
 $(t_j, \text{"SVM"})$   
 $\vdots$

Retrieval Function 1

$$f_1(x) \rightarrow y_1$$

Which one  
is better?

Retrieval Function 2

$$f_2(x) \rightarrow y_2$$

1. Kernel Machines  
<http://svm.first.gmd.de/>
2. SVM-Light Support Vector Machine  
<http://svmlight.joachims.org/>
3. School of Veterinary Medicine at UPenn  
<http://www.vet.upenn.edu/>
4. An Introduction to Support Vector Machines  
<http://www.support-vector.net/>
5. Service Master Company  
<http://www.servicemaster.com/>

$\vdots$

$U(t_j, \text{"SVM"}, y_1)$

1. School of Veterinary Medicine at UPenn  
<http://www.vet.upenn.edu/>
2. Service Master Company  
<http://www.servicemaster.com/>
3. Support Vector Machine  
<http://jbolivar.freesevers.com/>
4. Archives of SUPPORT-VECTOR-MACHINES  
<http://www.jiscmail.ac.uk/lists/SUPPORT...>
5. SVM-Light Support Vector Machine  
[http://ais.gmd.de/~thorsten/svm light/](http://ais.gmd.de/~thorsten/svm%20light/)

$\vdots$

$U(t_j, \text{"SVM"}, y_2)$

# Measuring Utility

Name	Description	Aggregation	Hypothesized Change with Decreased Quality
Abandonment Rate	% of queries with no click	N/A	Increase
Reformulation Rate	% of queries that are followed by reformulation	N/A	Increase
Queries per Session	Session = no interruption of more than 30 minutes	Mean	Increase
Clicks per Query	Number of clicks	Mean	Decrease
Click@1	% of queries with clicks at position 1	N/A	Decrease
Max Reciprocal Rank*	1/rank for highest click	Mean	Decrease
Mean Reciprocal Rank*	Mean of 1/rank for all clicks	Mean	Decrease
Time to First Click*	Seconds before first click	Median	Increase
Time to Last Click*	Seconds before final click	Median	Decrease

(\*) only queries with at least one click count

# ArXiv.org: User Study

## User Study in ArXiv.org

- Natural user and query population
- User in natural context, not lab
- Live and operational search engine
- Ground truth by construction

ORIG  $\succ$  SWAP2  $\succ$  SWAP4

- ORIG: Hand-tuned fielded
- SWAP2: ORIG with 2 pairs swapped
- SWAP4: ORIG with 4 pairs swapped

ORIG  $\succ$  FLAT  $\succ$  RAND

- ORIG: Hand-tuned fielded
- FLAT: No field weights
- RAND : Top 10 of FLAT shuffled

arXiv.org Full Text Search Results

Displaying hits 1 to 10 of 622. [Reorder by date.](#)

- [Emmanuel Monfni, Yann Guemour, A Quadratic Loss Multi-Class SVM \(2008\)](#)  
abstract: ... on the leave-one-out error of the pattern recognition SVM have been derived. Among these bounds, the most popular one ... bound. It applies to the hard margin pattern recognition SVM, and by extension to the 2-norm SVM. In this report, we introduce a quadratic loss M-SVM, the M-SVM<sup>2</sup>, as a di ...  
<http://arxiv.org/abs/0804.4898>
- [Nathalie Villa, Fabrice Rossi, Un résultat de consistance pour des SVM fonctionnels par interpolation spline \(2007\)](#)  
abstract: ... for function classification with Support Vector Machine (SVM). Rather than relying on projection on a truncated ... an implicit spline interpolation that allows us to compute SVM on the derivatives of the studied functions. To that end, w ...  
<http://arxiv.org/abs/0705.0210>
- [François Rapoport, Emmanuel Berthet and Jean-Philippe Vert, Classification of arrayCGH data using a fused SVM \(2008\)](#)  
abstract: ... a new method for supervised classification of arrayCGH data. The method is a variant of support vector machine (SVM) that incorporates the biological specificities of DNA copy number variations along the genome as prior knowledge. The ...  
<http://arxiv.org/abs/0801.3007>
- [Seung-ho Wu, Hui Zou, Ming Yuan, Structure variable selection in support vector machines \(2007\)](#)  
abstract: When applying the support vector machine (SVM) to high-dimensional classification problems, we often impose a sparse structure in the SVM to eliminate the influences of the irrelevant predictors. ... selection techniques have been successfully used in the SVM to perform automatic variable selection ...  
<http://arxiv.org/abs/0710.0508>
- [Marco Frullis, Oriana Mansutti, Praveen Boinep et al., A third level trigger programmable on FPGA for the gamma/hadron separation in a Cherenkov telescope using pseudo-Zernike moments and the SVM classifier \(2005\)](#)  
abstract: ... computed Pseudo-Zernike features as classification parameters. We implemented on a FPGA board a kernel function of the SVM and the Pseudo-Zernike features to build a third level trigger for the gamma-hadron separation task of the MAGIC Expen ...  
<http://arxiv.org/abs/cs/0602083>
- [Hao Helen Zhang, Yufeng Liu, Yichao Wu et al., Variable selection for the multiclass SVM via adaptive sup-norm regularization \(2008\)](#)  
abstract: The Support Vector Machine (SVM) is a popular classification paradigm in machine learning. ... great success in real applications. However, the standard SVM can not select variables ... of regularization in the context of the multiclass SVM (MSVM) for simultaneous classification and variable sel ...  
<http://arxiv.org/abs/0803.3676>
- [Seung-chan Ahn, Gene Kim and MyungHo Kim, A Note on Applications of Support Vector Machine \(2001\)](#)  
abstract: We describe in a rudimentary fashion how SVM (support vector machine) plays the role of classifier in a mathematical setting. We then discuss its application in the ...  
<http://arxiv.org/abs/math/0105169>
- [Haoshen Li, J. W. Clark, E. Mavrommatis et al., Modeling Nuclear Properties with Support Vector Machines \(2005\)](#)  
abstract: ... studies of the potential of support vector machines (SVM) for providing statistical models of nuclear systematics with demonstrable predictive power. Using SVM regression and classification procedures, we have created ...  
<http://arxiv.org/abs/nuc-th/0506080>
- [Gilles Blanchard, Olivier Bousquet, Pascal Massart, Statistical performance of support vector machines \(2008\)](#)  
abstract: ... studies of the potential of support vector machines (SVM) for providing statistical models of nuclear systematics with demonstrable predictive power. Using SVM regression and classification procedures, we have created ... compare to the penalty actually used in the SVM algorithm; (2) is ...  
<http://arxiv.org/abs/0804.0931>
- [Emidio Capriotti and Rita Casadio, The evaluation of protein folding rate constant is improved by predicting the folding kinetic order with a SVM-based method \(2006\)](#)  
abstract: ... first we describe a support vector machine-based method (SVM-KO) to predict for a given protein the kinetic order of the ... value can be obtained as a linear regression task with a SVM-based method. In this paper we show that linear correlation ...  
<http://arxiv.org/abs/q-bio.BM/0602013>

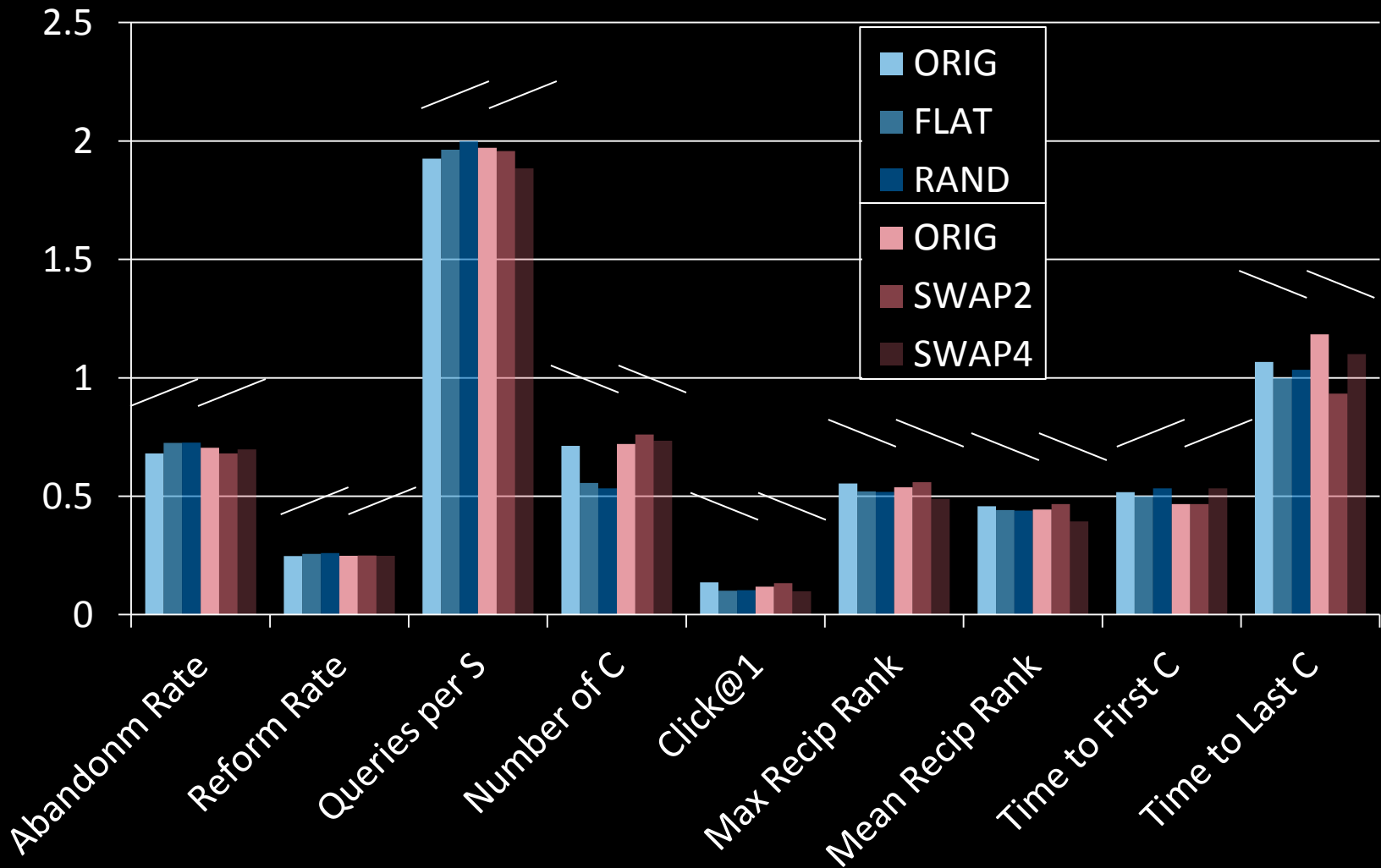
[Next >>](#)

# ArXiv.org: Experiment Setup

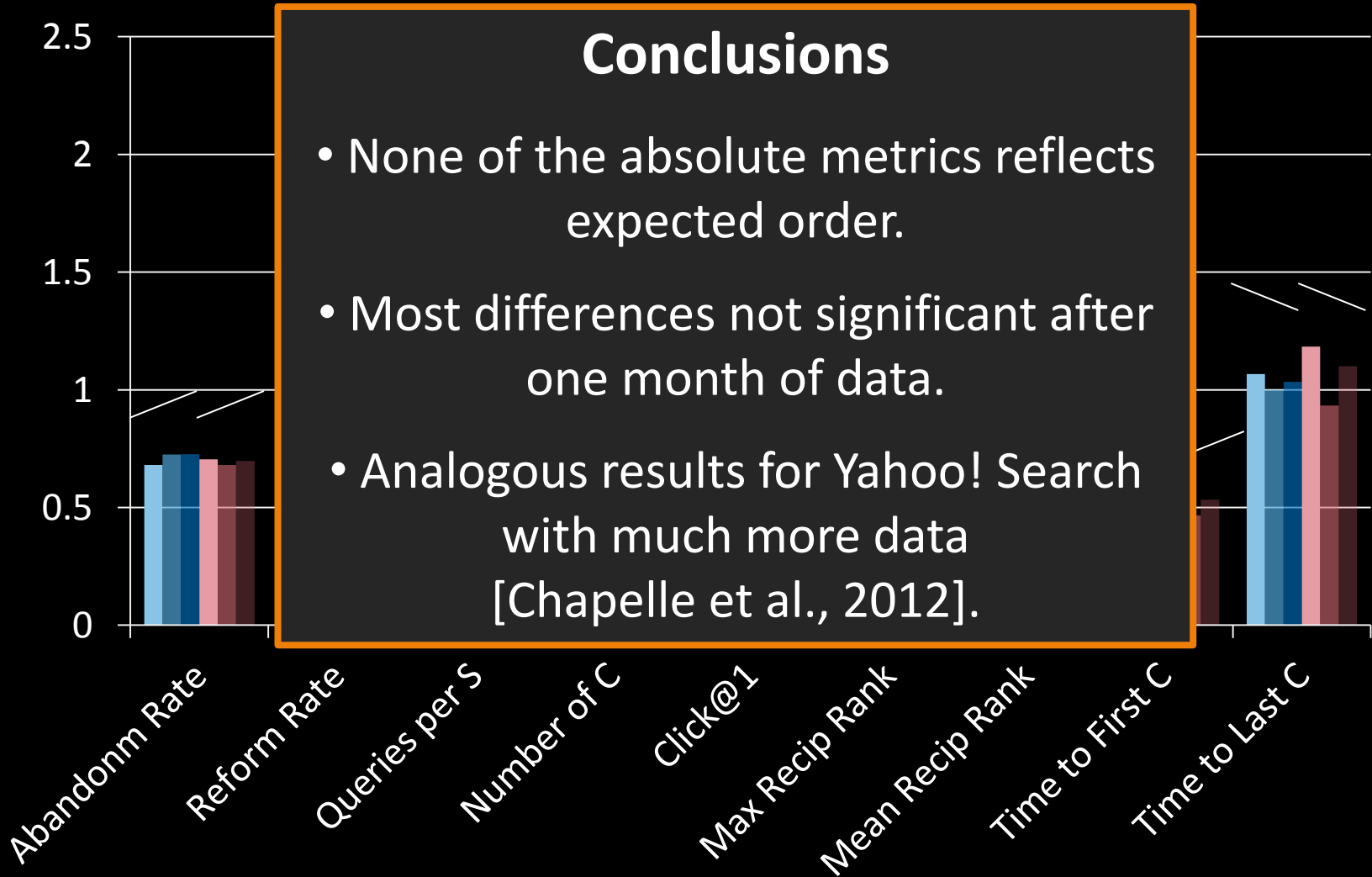
- Experiment Setup
  - Phase I: 36 days
    - Users randomly receive ranking from Orig, Flat, Rand
  - Phase II: 30 days
    - Users randomly receive ranking from Orig, Swap2, Swap4
  - User are permanently assigned to one experimental condition based on IP address and browser.
- Basic Statistics
  - ~700 queries per day / ~300 distinct users per day
- Quality Control and Data Cleaning
  - Test run for 32 days
  - Heuristics to identify bots and spammers
  - All evaluation code was written twice and cross-validated



# Arxiv.org: Results



# Arxiv.org: Results



# How about Explicit Feedback?

35 of 36 people found the following review helpful

★★★★☆ **a great buy for the price, a little hard to navigate for books**, November 13, 2012

By [flipnmelo](#) - [See all my reviews](#)

**This review is from:** [Kindle, 6" E Ink Display, Wi-Fi - for international shipment \(Electronics\)](#)

I purchased this for my fiance, and it's great overall, considering we live in a place with no bookstore and having to pay for shipping costs every time we wanted to read something was getting too much. With the Kindle we can download a book in no time.

What I don't like is that when searching for books under certain categories, there is no way to navigate to the end of the list any quicker. If I purchased all the books under the list on the first 5 pages, lets say, of comedy, and I'm ready to buy another I still have to manually go through all the pages of the list again, there's no way to skip ahead. But if we know what book we want, then it's worth it.

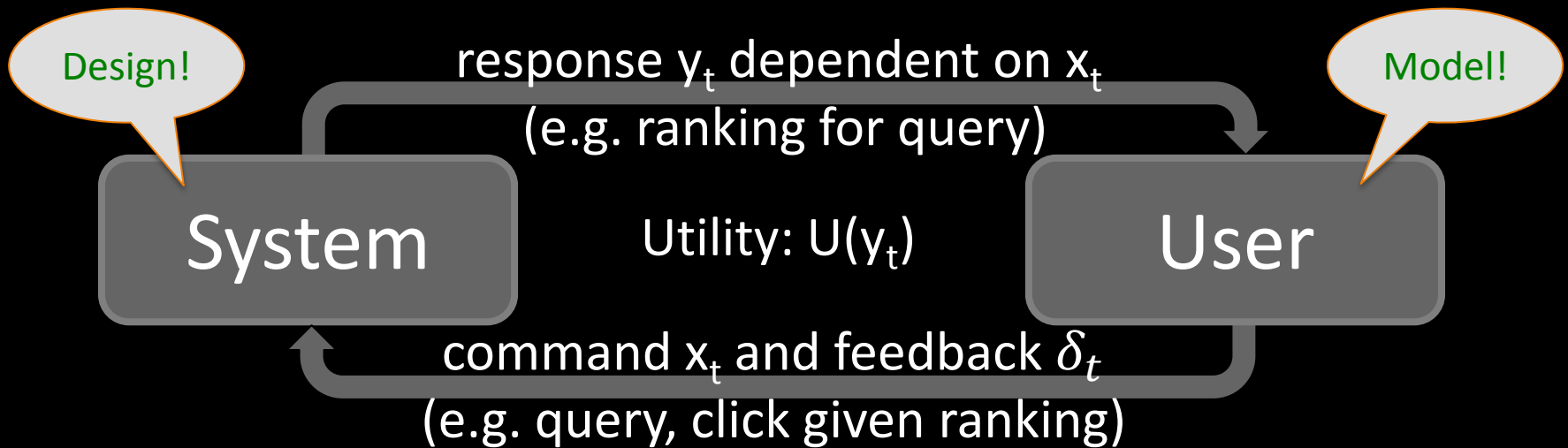
Also, the cost of some of the downloads are just as expensive as buying the books themselves. I thought that maybe it would cost less since there would be no costs for the print, paper, ink etc.

Help other customers find the most helpful reviews

Was this review helpful to you?

[Report abuse](#) | [Permalink](#)

# Interactive Learning System



- Observed Data  $\neq$  Training Data ✓
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Model the users decision process to extract feedback
  - Design learning algorithm for this type of feedback

# Decide between two Ranking Functions

Distribution  $P(x)$   
of  $x=(\text{user, query})$

$\vdots$   
 $(t_j, \text{"SVM"})$   
 $\vdots$

Retrieval Function 1

$$f_1(x) \rightarrow y_1$$

Which one  
is better?

Retrieval Function 2

$$f_2(x) \rightarrow y_2$$

1. Kernel Machines  
<http://svm.first.gmd.de/>
2. SVM-Light Support Vector Machine  
<http://svmlight.joachims.org/>
3. School of Veterinary Medicine at UPenn  
<http://www.vet.upenn.edu/>
4. An Introduction to Support Vector Machines  
<http://www.support-vector.net/>
5. Service Master Company  
<http://www.servicemaster.com/>

$\vdots$

$U(t_j, \text{"SVM"}, y_1)$

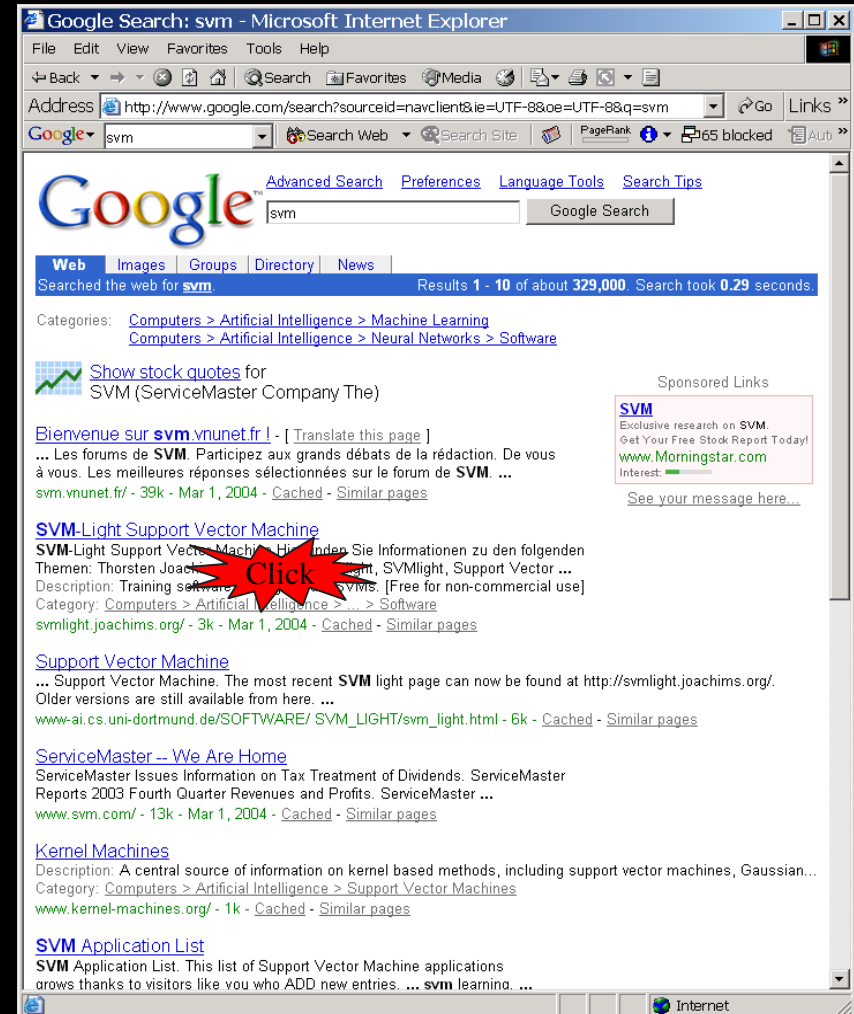
1. School of Veterinary Medicine at UPenn  
<http://www.vet.upenn.edu/>
2. Service Master Company  
<http://www.servicemaster.com/>
3. Support Vector Machine  
<http://jbolivar.freesevers.com/>
4. Archives of SUPPORT-VECTOR-MACHINES  
<http://www.jiscmail.ac.uk/lists/SUPPORT...>
5. SVM-Light Support Vector Machine  
[http://ais.gmd.de/~thorsten/svm light/](http://ais.gmd.de/~thorsten/svm%20light/)

$\vdots$

$U(t_j, \text{"SVM"}, y_2)$

# Economic Models of Decision Making

- Rational Choice
  - Alternatives  $\mathcal{Y}$
  - Utility function  $U(y)$
  - Decision  $y^* = \operatorname{argmax}_{y \in \mathcal{Y}} \{U(y)\}$
- Bounded Rationality
  - Time constraints
  - Computation constraints
  - Approximate  $U(y)$
- Behavioral Economics
  - Framing
  - Fairness
  - Loss aversion
  - Handling uncertainty



# A Model of how Users Click in Search

- Model of clicking:
  - Users explore ranking to position  $k$
  - Users click on most relevant (looking) links in top  $k$
  - Users stop clicking when time budget up or other action more promising (e.g. reformulation)
  - Empirically supported by [Granka et al., 2004]

Google Search: svm - Microsoft Internet Explorer

Address: <http://www.google.com/search?sourceid=navclient&ie=UTF-8&oe=UTF-8&q=svm>

Google Search

Searched the web for **svm**. Results 1 - 10 of about 329,000. Search took 0.29 seconds.

Categories: [Computers > Artificial Intelligence > Machine Learning](#)  
[Computers > Artificial Intelligence > Neural Networks > Software](#)

[Show stock quotes for SVM \(ServiceMaster Company The\)](#)

[Bienvenue sur svm.vnunet.fr !](#) - [ Translate this page ]  
... Les forums de SVM. Participez aux grands débats de la rédaction. De vous à vous. Les meilleures réponses sélectionnées sur le forum de SVM. ...  
svm.vnunet.fr/ - 39k - Mar 1, 2004 - Cached - Similar pages

[SVM-Light Support Vector Machine](#)  
SVM-Light Support Vector Machine. Hier finden Sie Informationen zu den folgenden Themen: Thorsten Joachims, SVMlight, Support Vector ...  
Description: Training software for SVMs. [Free for non-commercial use]  
Category: [Computers > Artificial Intelligence > Software](#)  
svm.light.joachims.org/ - 3k - Mar 1, 2004 - Cached - Similar pages

[Support Vector Machine](#)  
... Support Vector Machine. The most recent SVM light page can now be found at <http://svmlight.joachims.org/>. Older versions are still available from here. ...  
www.ai.cs.uni-dortmund.de/SOFTWARE/SVM\_LIGHT/svm\_light.html - 6k - Cached - Similar pages

[ServiceMaster -- We Are Home](#)  
ServiceMaster Issues Information on Tax Treatment of Dividends. ServiceMaster Reports 2003 Fourth Quarter Revenues and Profits. ServiceMaster ...  
www.svm.com/ - 13k - Mar 1, 2004 - Cached - Similar pages

[Kernel Machines](#)  
Description: A central source of information on kernel based methods, including support vector machines, Gaussian ...  
Category: [Computers > Artificial Intelligence > Support Vector Machines](#)  
www.kernel-machines.org/ - 1k - Cached - Similar pages

[SVM Application List](#)  
SVM Application List. This list of Support Vector Machine applications grows thanks to visitors like you who ADD new entries. ... svm learning. ...

**Click**

$\operatorname{argmax}_{y \in \text{Top } k} U(y)$

# Balanced Interleaving

$$x=(u=tj, q="svm")$$

$$f_1(x) \rightarrow y_1$$

$$f_2(x) \rightarrow y_2$$

1. Kernel Machines  
<http://svm.first.gmd.de/>
2. Support Vector Machine  
<http://jbolivar.freesevers.com/>
3. An Introduction to Support Vector Machines  
<http://www.support-vector.net/>
4. Archives of SUPPORT-VECTOR-MACHINES ...  
<http://www.jiscmail.ac.uk/lists/SUPPORT...>
5. SVM-Light Support Vector Machine  
[http://ais.gmd.de/~thorsten/svm\\_light/](http://ais.gmd.de/~thorsten/svm_light/)

1. Kernel Machines  
<http://svm.first.gmd.de/>
2. SVM-Light Support Vector Machine  
[http://ais.gmd.de/~thorsten/svm\\_light/](http://ais.gmd.de/~thorsten/svm_light/)
3. Support Vector Machine and Kernel ... References  
<http://svm.research.bell-labs.com/SVMrefs.html>
4. Lucent Technologies: SVM demo applet  
<http://svm.research.bell-labs.com/SVT/SVMsvt.html>
5. Royal Holloway Support Vector Machine  
<http://svm.dcs.rhnc.ac.uk>

## Interleaving( $y_1, y_2$ )

- |    |   |   |
|----|---|---|
| 1. | Kernel Machines   | 1 |
|    | <a href="http://svm.first.gmd.de/">http://svm.first.gmd.de/</a>   |   |
| 2. | Support Vector Machine  | 2 |
|    | <a href="http://jbolivar.freesevers.com/">http://jbolivar.freesevers.com/</a>                                     |   |
| 3. | SVM-Light Support Vector Machine  | 2 |
|    | <a href="http://ais.gmd.de/~thorsten/svm_light/">http://ais.gmd.de/~thorsten/svm_light/</a>                       |   |
| 4. | An Introduction to Support Vector Machines  | 3 |
|    | <a href="http://www.support-vector.net/">http://www.support-vector.net/</a>                                       |   |
| 5. | Support Vector Machine and Kernel ... References  | 3 |
|    | <a href="http://svm.research.bell-labs.com/SVMrefs.html">http://svm.research.bell-labs.com/SVMrefs.html</a>       |   |
| 6. | Archives of SUPPORT-VECTOR-MACHINES ...   | 4 |
|    | <a href="http://www.jiscmail.ac.uk/lists/SUPPORT...">http://www.jiscmail.ac.uk/lists/SUPPORT...</a>               |   |
| 7. | Lucent Technologies: SVM demo applet  | 4 |
|    | <a href="http://svm.research.bell-labs.com/SVT/SVMsvt.html">http://svm.research.bell-labs.com/SVT/SVMsvt.html</a> |   |

### Model of User:

Better retrieval functions  
is more likely to get more  
clicks.

### Invariant:

For all  $k$ , top  $k$  of  
balanced interleaving is  
union of top  $k_1$  of  $r_1$  and  
top  $k_2$  of  $r_2$  with  $k_1=k_2 \pm 1$ .

**Interpretation:**  $(y_1 \succ y_2) \Leftrightarrow \text{clicks}(\text{topk}(y_1)) > \text{clicks}(\text{topk}(y_2))$

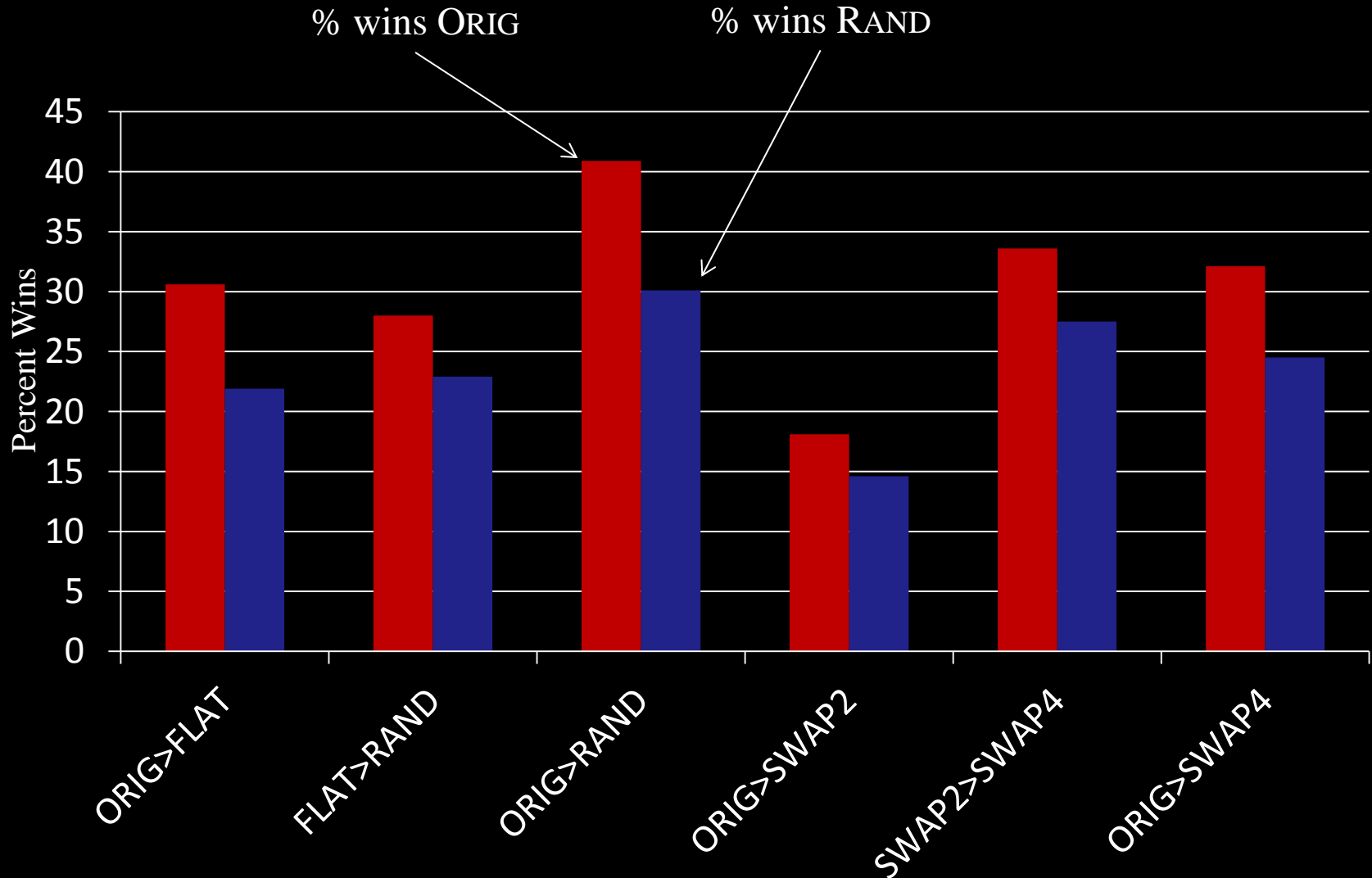
$\rightarrow$  see also [Radlinski, Craswell, 2012] [Hofmann, 2012]



# Arxiv.org: Interleaving Experiment

- Experiment Setup
  - Phase I: 36 days
    - Balanced Interleaving of (Orig,Flat) (Flat,Rand) (Orig,Rand)
  - Phase II: 30 days
    - Balanced Interleaving of (Orig,Swap2) (Swap2,Swap4) (Orig,Swap4)
- Quality Control and Data Cleaning
  - Same as for absolute metrics

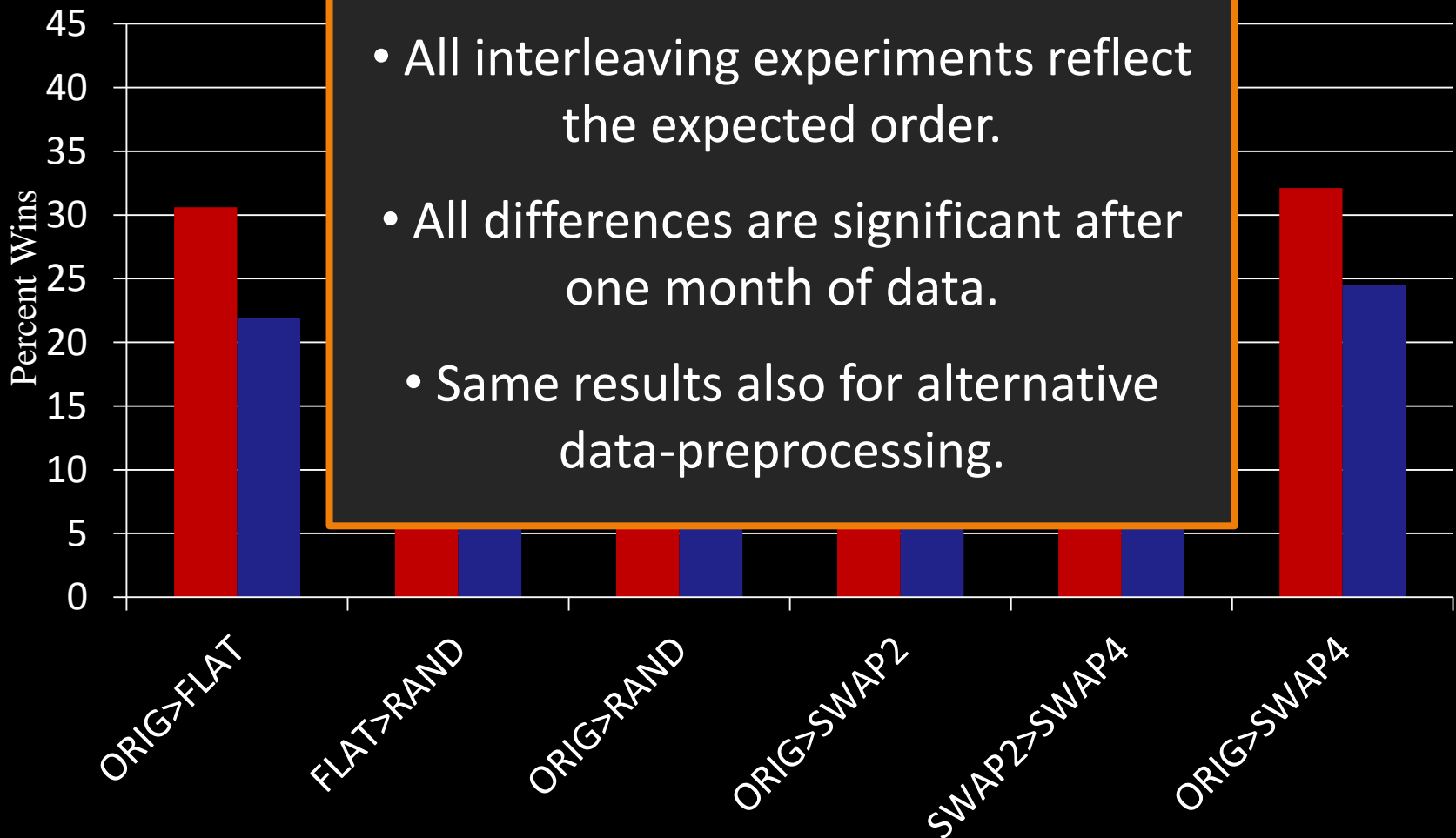
# Arxiv.org: Interleaving Results



# Arxiv.org: Interleaving Results

## Conclusions

- All interleaving experiments reflect the expected order.
- All differences are significant after one month of data.
- Same results also for alternative data-preprocessing.



# Yahoo and Bing: Interleaving Results

- Yahoo Web Search [Chapelle et al., 2012]
  - Four retrieval functions (i.e. 6 paired comparisons)
  - Balanced Interleaving
    - All paired comparisons consistent with ordering by NDCG.
- Bing Web Search [Radlinski & Craswell, 2010]
  - Five retrieval function pairs
  - Team-Game Interleaving
    - Consistent with ordering by NDGC when NDCG significant.

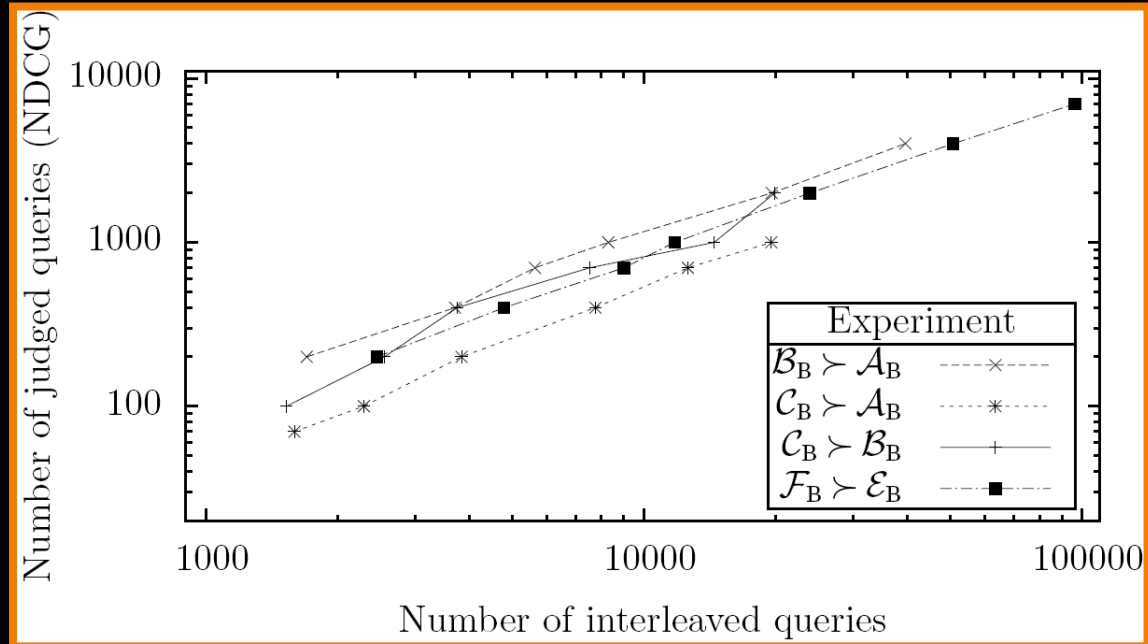
# Efficiency: Interleaving vs. Explicit

- Bing Web Search
  - 4 retrieval function pairs
  - ~12k manually judged queries
  - ~200k interleaved queries

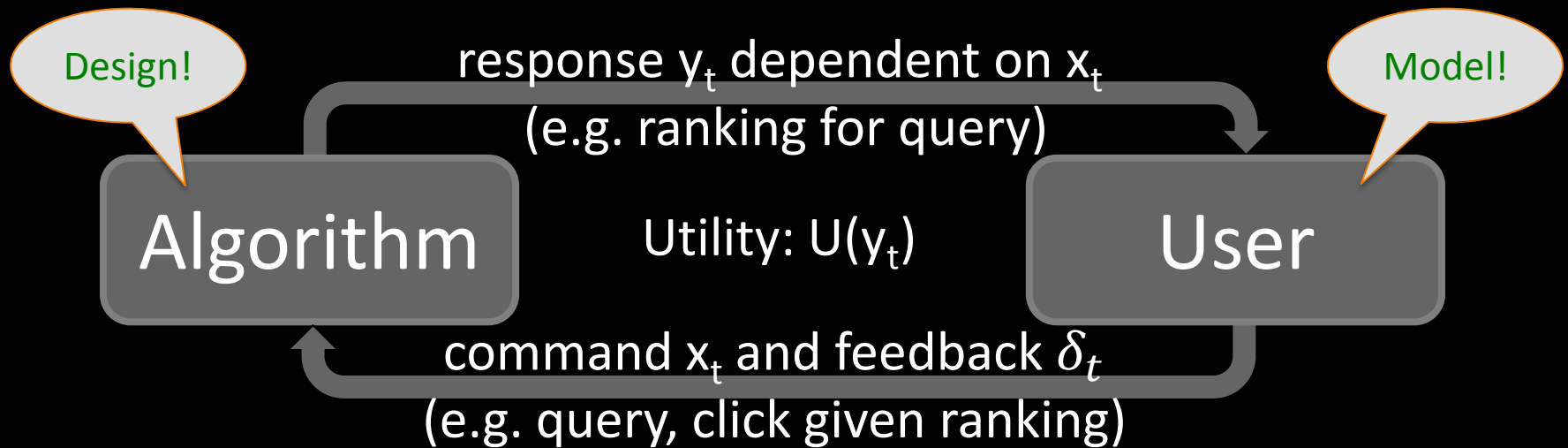
- Experiment

- $p$  = probability that NDCG is correct on subsample of size  $y$
- $x$  = number of queries needed to reach same  $p$ -value with interleaving

➔ Ten interleaved queries are equivalent to one manually judged query.



# Interactive Learning System



- Observed Data  $\neq$  Training Data ✓
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Model the users decision process to extract feedback ✓
    - $\rightarrow$  Pairwise comparison test  $P(y_i \succ y_j \mid U(y_i) > U(y_j))$
  - Design learning algorithm for this type of feedback

# Supervised Batch Learning

- Data:  $(x, y^*) \sim P(X, Y^*)$ 
    - $x$  is input
    - $y^*$  is true label
  - Rules:  $f \in H$
  - Prediction:  $\hat{y} = f(x)$
  - Loss function:  $\Delta(\hat{y}, y^*)$
- Find  $f \in H$  that minimizes prediction error

$$R(f) = \int \Delta(f(x), y^*)$$

# Learning on Operational System

- Example: 4 retrieval functions:  $A > B \gg C > D$ 
  - 10 possible pairs for interactive experiment
    - (A,B)  $\rightarrow$  low cost to user
    - (A,C)  $\rightarrow$  medium cost to user
    - (C,D)  $\rightarrow$  high cost to user
    - (A,A)  $\rightarrow$  zero cost to user
    - ...
- Minimizing Regret
  - Don't present "bad" pairs more often than necessary
  - Trade off (long term) informativeness and (short term) cost
  - Definition: Probability of  $(f_t, f'_t)$  losing against the best  $f^*$

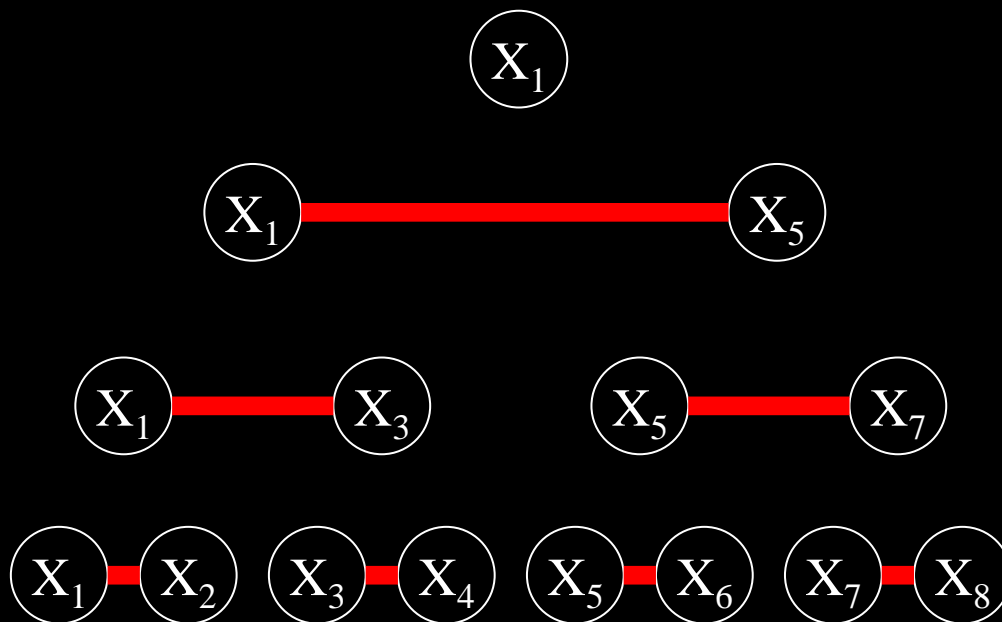
$$R(A) = \sum_{t=1}^T [P(f^* \succ f_t) - 0.5] + [P(f^* \succ f'_t) - 0.5]$$

**$\rightarrow$  Dueling Bandits Problem**



# First Thought: Tournament

- Noisy Sorting/Max Algorithms:
  - [Feige et al.]: Triangle Tournament Heap  $O(n/\varepsilon^2 \log(1/\delta))$  with prob  $1-\delta$
  - [Adler et al., Karp & Kleinberg]: optimal under weaker assumptions



# Algorithm: Interleaved Filter 2

- Algorithm

InterleavedFilter1( $T, W = \{f_1 \dots f_K\}$ )

- Pick random  $f'$  from  $W$
- $\delta = 1/(TK^2)$
- WHILE  $|W| > 1$ 
  - FOR  $b \in W$  DO
    - » duel( $f', b$ )
    - » update  $P_f$
  - $t = t + 1$
  - $c_t = (\log(1/\delta)/t)^{0.5}$
  - Remove all  $f$  from  $W$  with  $P_f < 0.5 - c_t$  [WORSE WITH PROB  $1 - \delta$ ]
  - IF there exists  $f''$  with  $P_{f''} > 0.5 + c_t$  [BETTER WITH PROB  $1 - \delta$ ]
    - » Remove  $f'$  from  $W$
    - » Remove all  $f$  from  $W$  that are empirically inferior to  $f'$
    - »  $f' = f''$ ;  $t = 0$
- UNTIL  $T$ : duel( $f', f'$ )

$f_1$	$f_2$	$f' = f_3$	$f_4$	$f_5$
0/0	0/0		0/0	0/0

$f_1$	$f_2$	$f' = f_3$	$f_4$	$f_5$
8/2	7/3		4/6	<del>1/9</del>

$f_1$	$f_2$	$f' = f_3$	$f_4$	
13/2	11/4	<del></del>	<del>7/8</del>	XX

$f' = f_1$	$f_2$		$f_4$	
0/0	0/0	XX	XX	XX

# Assumptions

- Preference Relation:  $f_i \succ f_j \Leftrightarrow P(f_i \succ f_j) = 0.5 + \varepsilon_{i,j} > 0.5$
- Weak Stochastic Transitivity:  $f_i \succ f_j$  and  $f_j \succ f_k \rightarrow f_i \succ f_k$

**Theorem:** IF2 incurs expected average regret bounded by

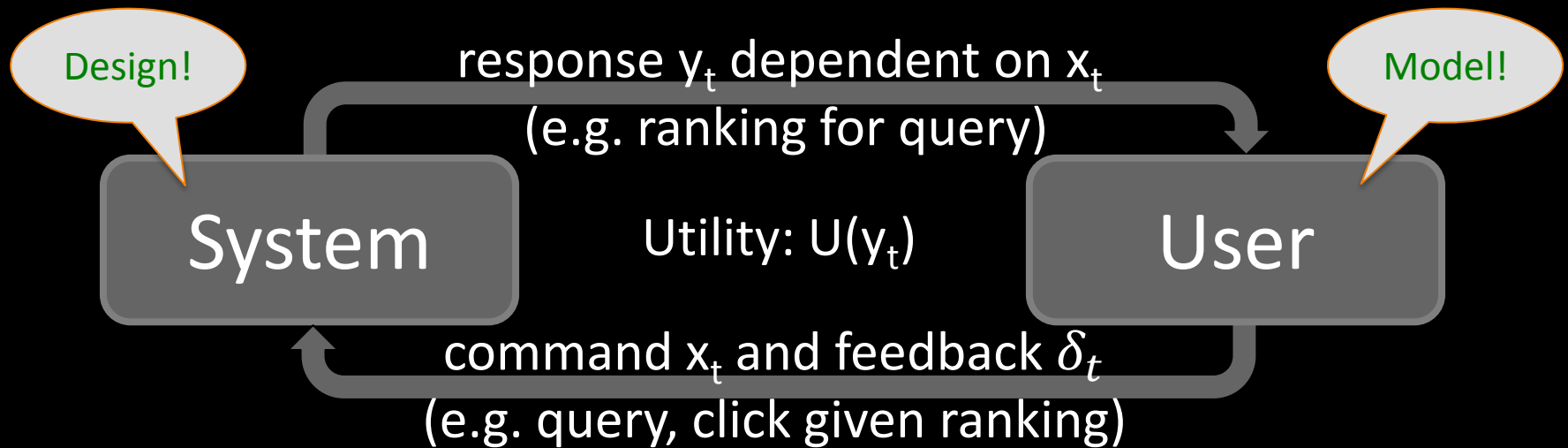
- $$\frac{1}{T} E(R_T) \leq O\left(\frac{K \log T}{\varepsilon_{1,2} T}\right)$$

- Stochastic Triangle Inequality:  $f_i \succ f_j \succ f_k \rightarrow \varepsilon_{i,k} \leq \varepsilon_{i,j} + \varepsilon_{j,k}$

$$\varepsilon_{1,2} = 0.01 \text{ and } \varepsilon_{2,3} = 0.01 \rightarrow \varepsilon_{1,3} \leq 0.02$$

- $\varepsilon$ -Winner exists:  $\varepsilon = \max_i \{ P(f_1 \succ f_i) - 0.5 \} = \varepsilon_{1,2} > 0$

# Interactive Learning System



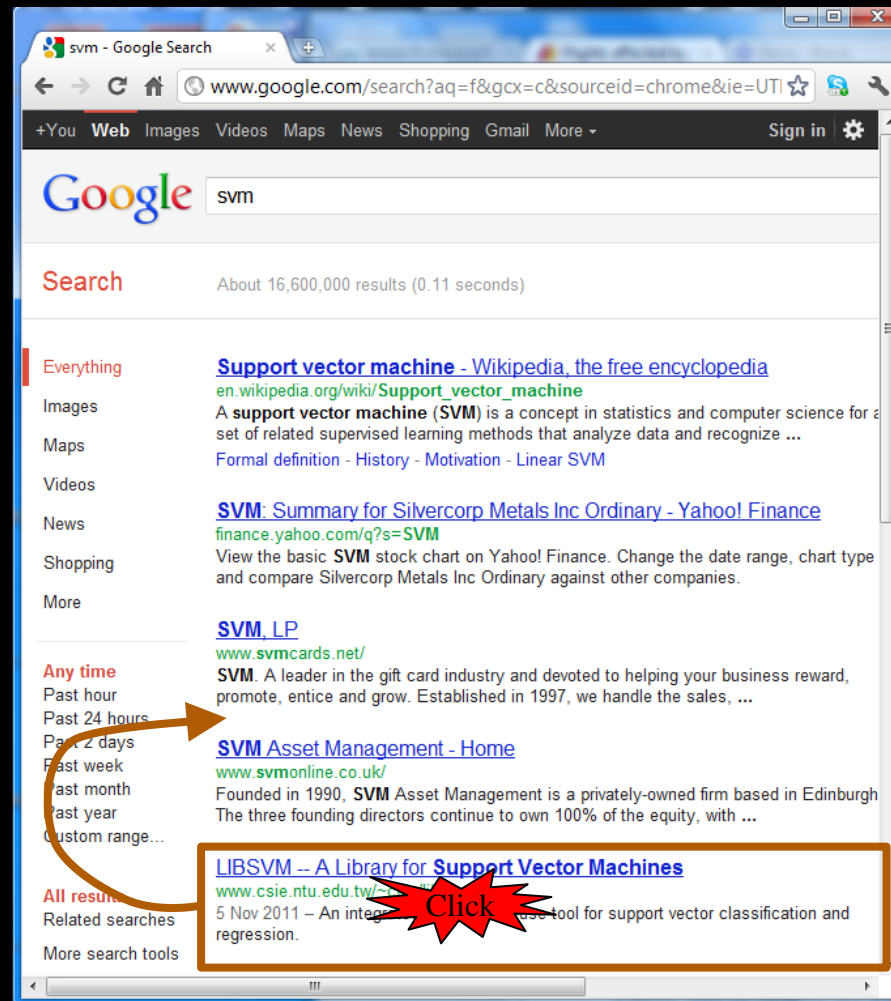
- Observed Data  $\neq$  Training Data ✓
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Model the users decision process to extract feedback ✓
    - $\rightarrow$  Pairwise comparison test  $P(y_i \succ y_j \mid U(y_i) > U(y_j))$
  - Design learning algorithm for this type of feedback ✓
    - $\rightarrow$  Dueling Bandits problem and algorithms (e.g. IF2)

# Who does the exploring? Example 1

The image shows a screenshot of the Netflix website interface. The main browser window displays the movie page for "Lie to Me" (URL: [movies.netflix.com/WiMovie/Lie\\_to\\_Me/70140406?trkid=13462049](http://movies.netflix.com/WiMovie/Lie_to_Me/70140406?trkid=13462049)). The page features a red navigation bar with "NETFLIX", "Watch Instantly", "Just for Kids", "Taste Profile", "DVDs", and "DVD Queue". A search bar contains the text "Movies, TV shows, actors, directors, genres" and a user profile icon for "Thorsten". Below the navigation bar is a pagination control showing "1" selected out of "75" items. The main content area is titled "More Like Lie to Me" and displays a row of movie posters: NUMB3RS, BONES, FLASHPOINT, AWAKE, CSI: NY, and KEEPER. A sidebar on the left shows "Recently Watched" (including "TALLER PARK BOYS" and "AZIZ ANSARI") and "Top Picks for Thorsten" (including "IN GOD WE TRUST" and "DO FRIES"). The footer contains copyright information for 2013 Netflix, Inc., and a "Service Code" section with a disclaimer: "Use of the Netflix service and this Web site constitutes acceptance of our Terms of Use and Privacy Policy. All rights reserved. About Cookies and Internet Advertising (1-ef8b98a)".

# Who does the exploring?

## Example 2



The image shows a screenshot of a Google search for "svm". The search results are displayed in a browser window. The search bar contains "svm" and the results show "About 16,600,000 results (0.11 seconds)".

The search results are categorized by type:

- Everything:** [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](#)
- News:** [SVM: Summary for Silvercorp Metals Inc Ordinary - Yahoo! Finance](#) (finance.yahoo.com/q?s=SVM). View the basic SVM stock chart on Yahoo! Finance. Change the date range, chart type and compare Silvercorp Metals Inc Ordinary against other companies.
- Shopping:** [SVM, LP](#) (www.svmcards.net/). 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, ...
- More:** [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 ...
- LIBSVM -- A Library for Support Vector Machines** (www.csie.ntu.edu.tw/~cjlin/libsvm/). 5 Nov 2011 - An integrated software tool for support vector classification and regression.

Annotations on the screenshot include:

- A red starburst with the word "Click" pointing to the LIBSVM link.
- An orange arrow pointing from the "Past 24 hours" filter to the SVM Asset Management link.
- An orange arrow pointing from the "Past 24 hours" filter to the LIBSVM link.
- An orange box around the LIBSVM link and its description.

# Who does the exploring? Example 3

The image displays two browser windows side-by-side, illustrating a search process. The left window shows a Google search for 'svm' with approximately 16,600,000 results. The right window shows a Google search for 'sv meppen' with approximately 939,000 results. An orange arrow points from the 'svm' search results to the 'sv meppen' search results, highlighting a specific result with a red starburst and the word 'Click'.

**Left Window: svm - Google Search**  
www.google.com/search?aq=f&gcx=c&sourceid=chrome&ie=UTF-8

Search: svm  
About 16,600,000 results (0.11 seconds)

**Everything**  
[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 that is used for classification and regression analysis.  
Formal definition - History - Motivation - Linear SVM

**Any time**  
Past hour  
Past 24 hours  
Past 2 days  
Past week  
Past month  
Past year  
Custom range...

**All results**  
Related searches  
More search tools

**Right Window: sv meppen - Google Search**  
www.google.com/search?aq=f&gcx=c&sourceid=chrome&ie=UTF-8

Search: sv meppen  
About 939,000 results (0.09 seconds)

**Everything**  
[SV Meppen 1912 e.V. - Offizielle Webseite](#)  
www.svmeppen.de/ - Translate this page  
Die offizielle Homepage des am 29. November 1912 gegründeten Fußballvereins präsentiert einen Live-Ticketverkauf und informiert über die Mannschaft.

**Willkommen auf www.svmeppen.de - SV Meppen 1912 e.V. ...**  
1912.svmeppen.de/ - Translate this page  
SV Meppen e.V. 1912 - Offizielle Website- ... SV Meppen, meppen, emsland, oberliga, oberliga nord, fussball, fußball, lingen, steve haensel, webcomtech.net, ...

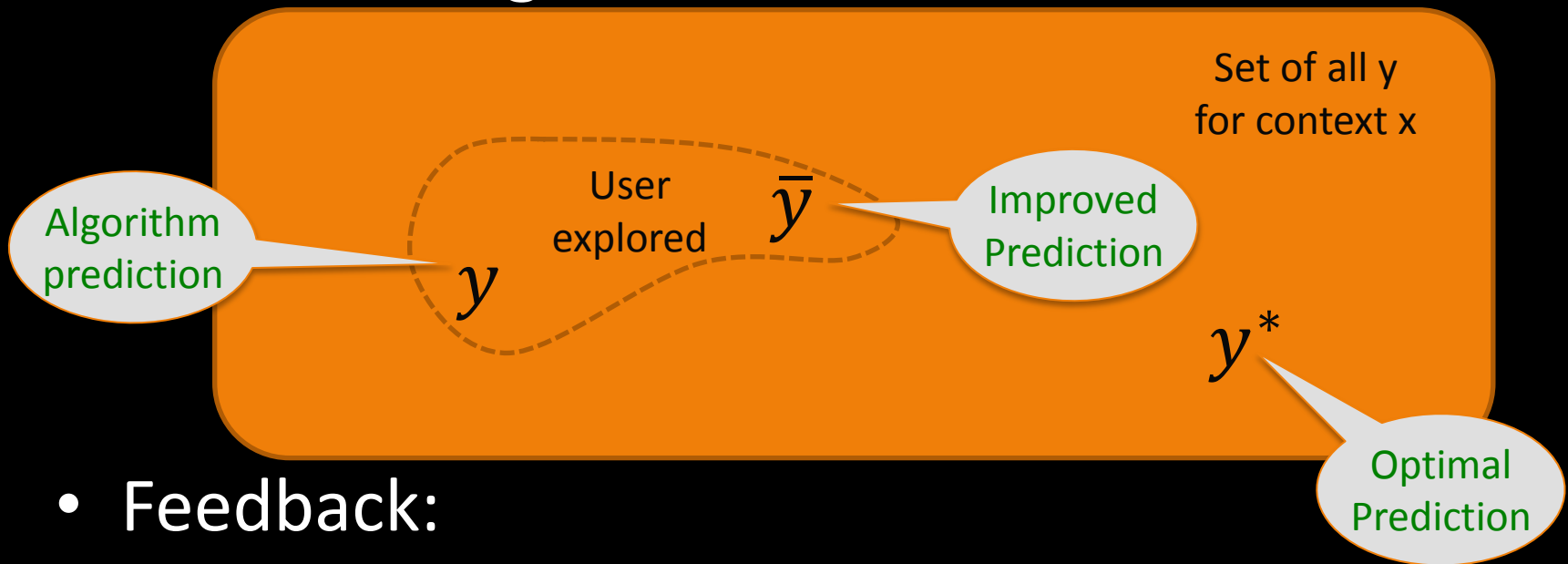
**SV Meppen - Wikipedia, the free encyclopedia**  
en.wikipedia.org/wiki/SV\_Meppen  
SV Meppen is a German association football club playing in Meppen, Lower Saxony. The club was founded on 29 November 1912 as Amisia Meppen and ...  
History - Stadium - Records - Literature

**SV Meppen - Nachrichten, Liveticker, Bilder vom SV Meppen in der ...**  
www.noz.de/sport/sv-meppen - Translate this page  
Berichte, Liveticker, Bilder und Audios vom SV Meppen, mehr zur Mannschaft sowie Analysen der Gegner in der Fußball-Regionalliga.

**SV Meppen - Fußballverein - transfermarkt.de**  
www.transfermarkt.de/.../sv-meppen/.../verein\_24... - Translate this page  
Mit dieser Nachricht hatte Stephen Famewo (Foto) nicht gerechnet. Als unumstrittener Stammspieler trug er dazu bei, dass der SV Meppen in die Regionalliga ...

# Coactive Feedback Model

- Interaction: given  $x$



- Feedback:

– Improved prediction  $\bar{y}_t$

$$U(\bar{y}_t | x_t) > U(y_t | x_t)$$

– Supervised learning: optimal prediction  $y_t^*$

$$y_t^* = \operatorname{argmax}_y U(y | x_t)$$



# Machine Translation

$x_t$

We propose Coactive Learning as a model of interaction between a learning system and a human user, where both have the common goal of providing results of maximum utility to the user.

$y_t$

Wir schlagen vor, koaktive Learning als ein Modell der Wechselwirkung zwischen einem Lernsystem und menschlichen Benutzer, wobei sowohl die gemeinsame Ziel, die Ergebnisse der maximalen Nutzen für den Benutzer.



Wir schlagen ~~vor~~, koaktive Learning als ein Modell ~~der Wechselwirkung des Dialogs~~ zwischen einem Lernsystem und menschlichen Benutzer, wobei ~~sowohl die beide das~~ gemeinsame Ziel ~~haben~~, die Ergebnisse der maximalen Nutzen für den Benutzer ~~zu liefern~~.

$\bar{y}_t$

# Supervised Batch Learning

- Data:  $(x, y^*) \sim P(X, Y^*)$ 
    - $x$  is input
    - $y^*$  is true label
  - Rules:  $f \in H$
  - Prediction:  $\hat{y} = f(x)$
  - Loss function:  $\Delta(\hat{y}, y^*)$
- Find  $f \in H$  that minimizes prediction error

$$R(f) = \int \Delta(f(x), y^*)$$

# Coactive Preference Perceptron

- Model
  - Linear model of user utility:  $U(y|x) = w^T \phi(x,y)$
- Algorithm
  - FOR  $t = 1$  TO  $T$  DO
    - Observe  $x_t$
    - Present  $y_t = \operatorname{argmax}_y \{ w_t^T \phi(x_t, y) \}$
    - Obtain feedback  $\bar{y}_t$  from user
    - Update  $w_{t+1} = w_t + \phi(x_t, \bar{y}_t) - \phi(x_t, y_t)$
- This may look similar to a multi-class Perceptron, but
  - Feedback  $\bar{y}_t$  is different (not get the correct class label)
  - Regret is different (misclassifications vs. utility difference)

$$R(A) = \frac{1}{T} \sum_{t=1}^T [U(y_t^*|x) - U(y_t|x)]$$

Never revealed:  
• cardinal feedback  
• optimal  $y^*$

# Coactive Perceptron: Regret Bound

- Model

$U(\mathbf{y}|\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}, \mathbf{y})$ , where  $w$  is unknown

- Feedback:  $\xi$ -Approximately  $\alpha$ -Informative

$$E[U(x_t, \bar{y}_t)] \geq U(x_t, y_t) + \alpha(U(x_t, y_t^*) - U(x_t, y_t)) - \xi_t$$

- Theorem

user  
feedback

system  
prediction

gap to  
optimal

model  
error

For user feedback  $\bar{y}$  that is  $\alpha$ -informative in expectation, the expected average regret of the Preference Perceptron is bounded by

$$E \left[ \frac{1}{T} \sum_{t=1}^T U(y_t^* | x) - U(y_t | x) \right] \leq \frac{1}{\alpha T} \sum_{t=1}^T \xi_t + \frac{2R ||w||}{\alpha \sqrt{T}}$$

→ zero

model error

# Preference Perceptron: Experiment

## Experiment:

- Automatically optimize Arxiv.org Fulltext Search

Analogous  
to DCG

## Model

- Utility of ranking  $y$  for query  $x$ :  $U_t(y|x) = \sum_i \gamma_i w_t^\top \phi(x, y^{(i)})$  [ $\sim 1000$  features]  
→ Computing argmax ranking: sort by  $w_t^\top \phi(x, y^{(i)})$

## Feedback

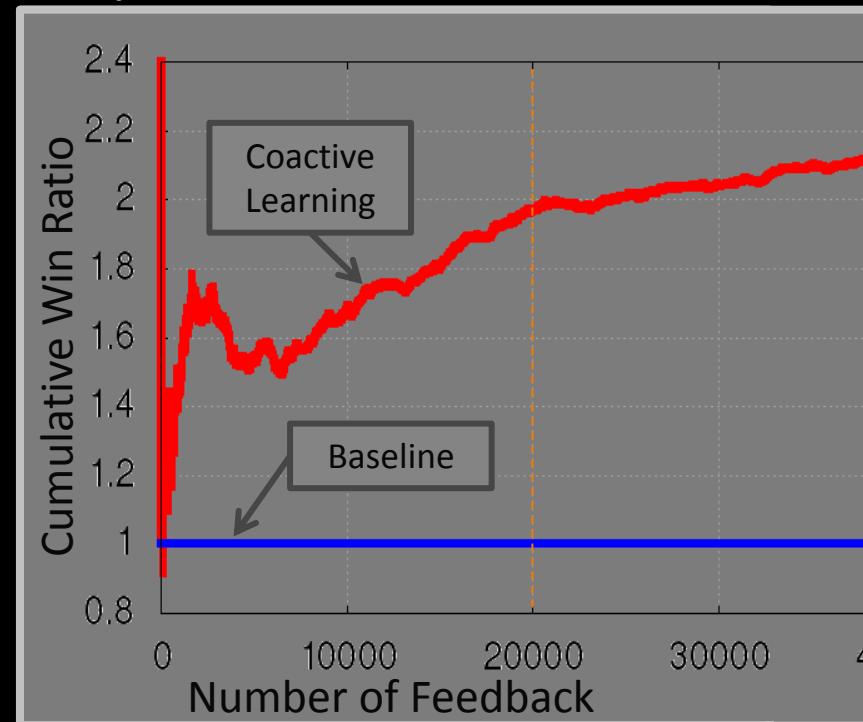
- Construct  $\bar{y}_t$  from  $y_t$  by moving clicked links one position higher.
- Perturbation [Raman et al., 2013]

## Baseline

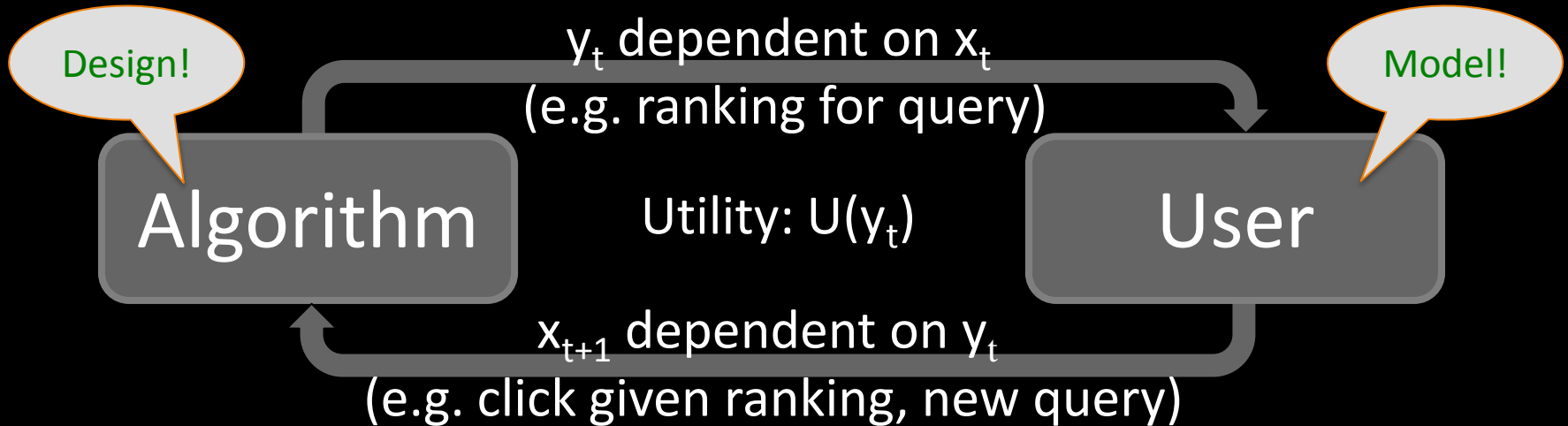
- Handtuned  $w_{\text{base}}$  for  $U_{\text{base}}(y|x)$

## Evaluation

- Interleaving of ranking from  $U_t(y|x)$  and  $U_{\text{base}}(y|x)$



# Interactive Learning System



- Observed Data  $\neq$  Training Data
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Dueling Bandits
    - $\rightarrow$  Model: Pairwise comparison test  $P(y_i \succ y_j \mid U(y_i) > U(y_j))$
    - $\rightarrow$  Algorithm: Interleaved Filter 2,  $O(|Y| \log(T))$  regret
  - Coactive Learning
    - $\rightarrow$  Model: for given  $y$ , user provides  $\bar{y}$  with  $U(\bar{y} \mid x) > U(y \mid x)$
    - $\rightarrow$  Algorithm: Preference Perceptron,  $O(\|w\| T^{0.5})$  regret

# Running Interactive Learning Experiments

~~1) Build your own system and provide service~~

~~→ not your thing~~

~~→ too little data~~

~~2) Convince others to run your experiments on  
commercial system~~

~~→ good luck with that~~

3) Use large-scale historical log data from  
commercial system

# Information in Interaction Logs

- Partial Information (aka “Bandit”) Feedback
  - News recommender  $f_0$  presents set  $y$  of articles for user  $x$  and observes that user reads  $\delta$  minutes
  - Ad system  $f_0$  presents ad  $y$  to user  $x$  and observes monetary payoff  $\delta$
  - Search engine  $f_0$  interleaves ranking  $y$  for query  $x$  with baseline ranker and observes win/loss  $\delta$
  - MT system  $f_0$  predicts translation  $y$  for  $x$  and receives rating  $\delta$

→ Data:  $S = ((x_1, y_1, \delta_1), \dots, (x_n, y_n, \delta_n))$

context

$f_0$  action

reward



# Supervised Batch Learning

- Data:  $(x, y^*) \sim P(X, Y^*)$ 
    - $x$  is input
    - $y^*$  is true label
  - Rules:  $f \in H$
  - Prediction:  $\hat{y} = f(x)$
  - Loss function:  $\Delta(\hat{y}, y^*)$
- Find  $f \in H$  that minimizes prediction error

$$R(f) = \int \Delta(f(x), y^*)$$

# Changing History

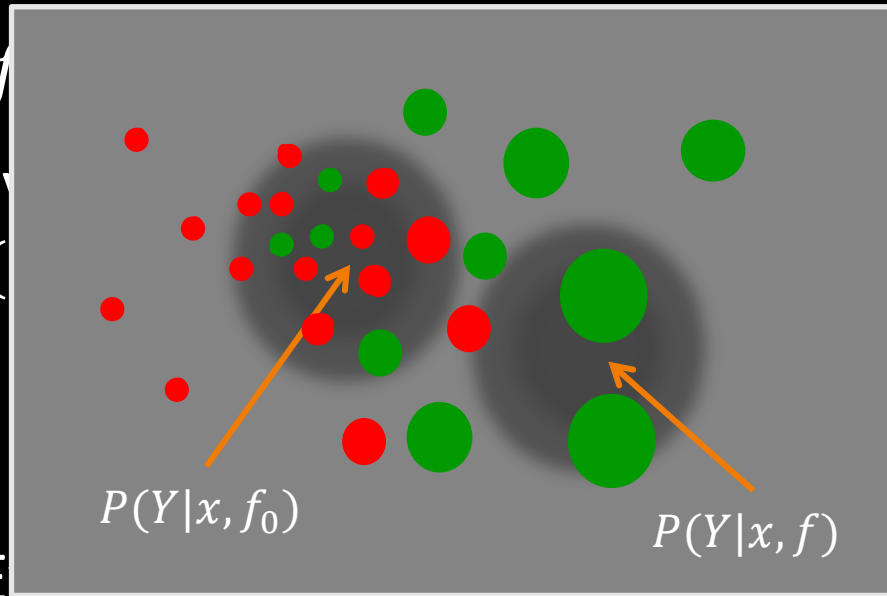
- Expected Performance of Stochastic Policy  $f$ :  $P(y|x, f)$

$R(f)$

- On-Policy Evaluation

Given  $S = (\dots)$

collected under  $f_0$ ,



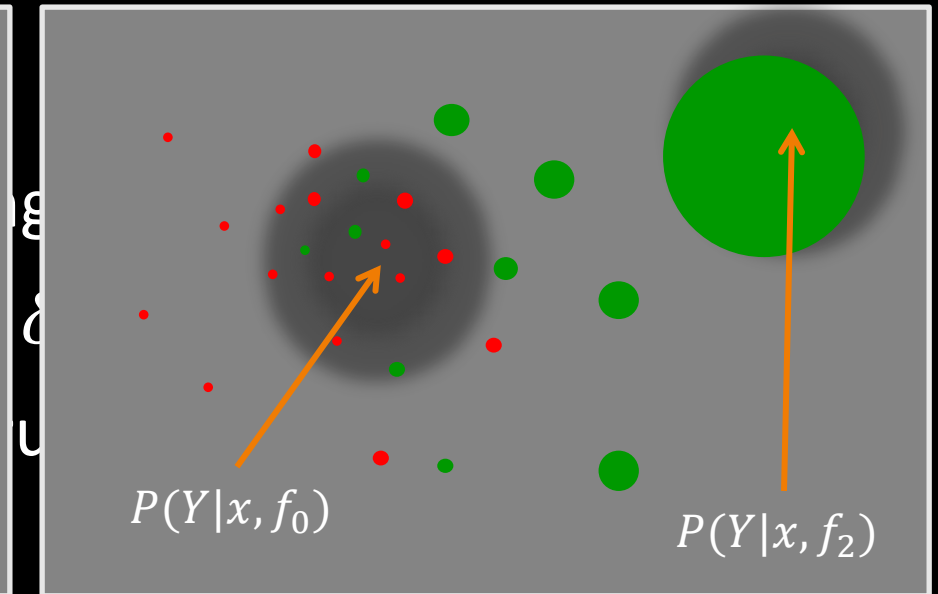
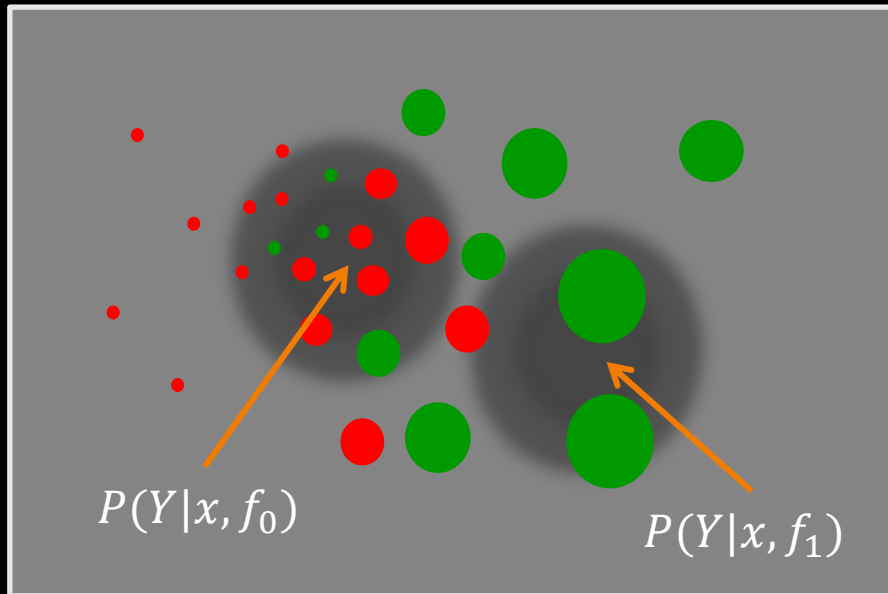
- Off-Policy Evaluation

Given  $S = ((x_1, y_1, \delta_1), \dots, (x_n, y_n, \delta_n))$  collected under  $f_0$ ,

$$\hat{R}(f) = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{P(y_i|x_i, f)}{P(y_i|x_i, f_0)}$$

Propensity weight

# Partial Information Empirical Risk Minimization



$$\hat{f} := \operatorname{argmax}_{f \in H} \sum_i^n \frac{P(y_i | x_i, h)}{p_i} \delta_i$$

# Partial Information Empirical Risk Minimization

- Setup

- Stochastic logging using  $f_0$  with  $p_i = P(y_i|x_i, f_0)$ 
  - Data  $S = ((x_1, y_1, \delta_1, p_1), \dots, (x_n, y_n, \delta_n, p_n))$
- Stochastic prediction rules  $f \in H: P(y_i|x_i, f)$

- Training

$$\hat{f} := \operatorname{argmax}_{f \in H} \sum_i^n \frac{P(y_i|x_i, h)}{p_i} \delta_i$$

# Propensity Risk Minimization

- Intuition and Learning Theory
  - De-bias estimator through propensity weighting
  - Correct for variance of estimators for different  $f \in H$
  - Account for capacity of the hypothesis space  $H$
- Training: optimize generalization error bound

$$\hat{f} = \operatorname{argmax}_{f \in H} \left[ \begin{array}{l} \widehat{\operatorname{Mean}} \left( \frac{P(y_i|x_i,f)}{p_i} \delta_i \right) \\ - \lambda_1 \sqrt{\widehat{\operatorname{Var}} \left( \frac{P(y_i|x_i,f)}{p_i} \delta_i \right)} \\ - \lambda_2 \operatorname{Reg}(H) \end{array} \right]$$

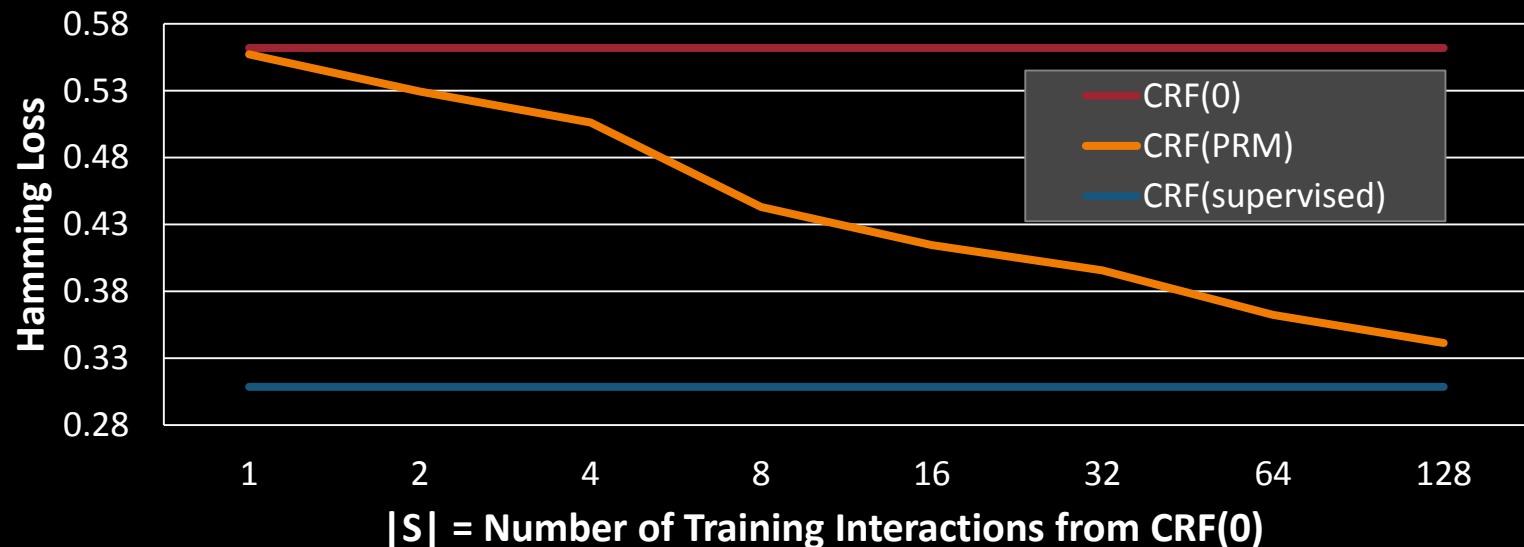
Unbiased Estimator

Variance Control

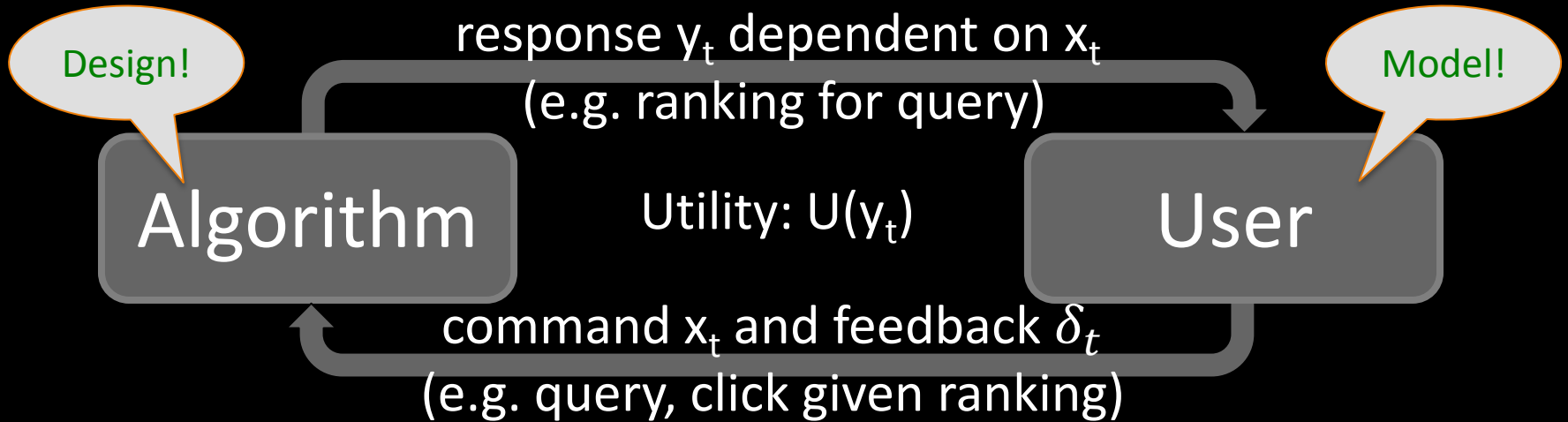
Capacity Control

# Experiment: Propensity Risk Minimization

- Experiment Setup
  - $x$ : Reuters RCV1 text document
  - $y$ : label vector with 4 binary labels
  - $\delta$ : number of correct labels
  - $p$ : propensity under logging with CRF(0)
  - $H$ : CRF with one weight vector per label
- Results

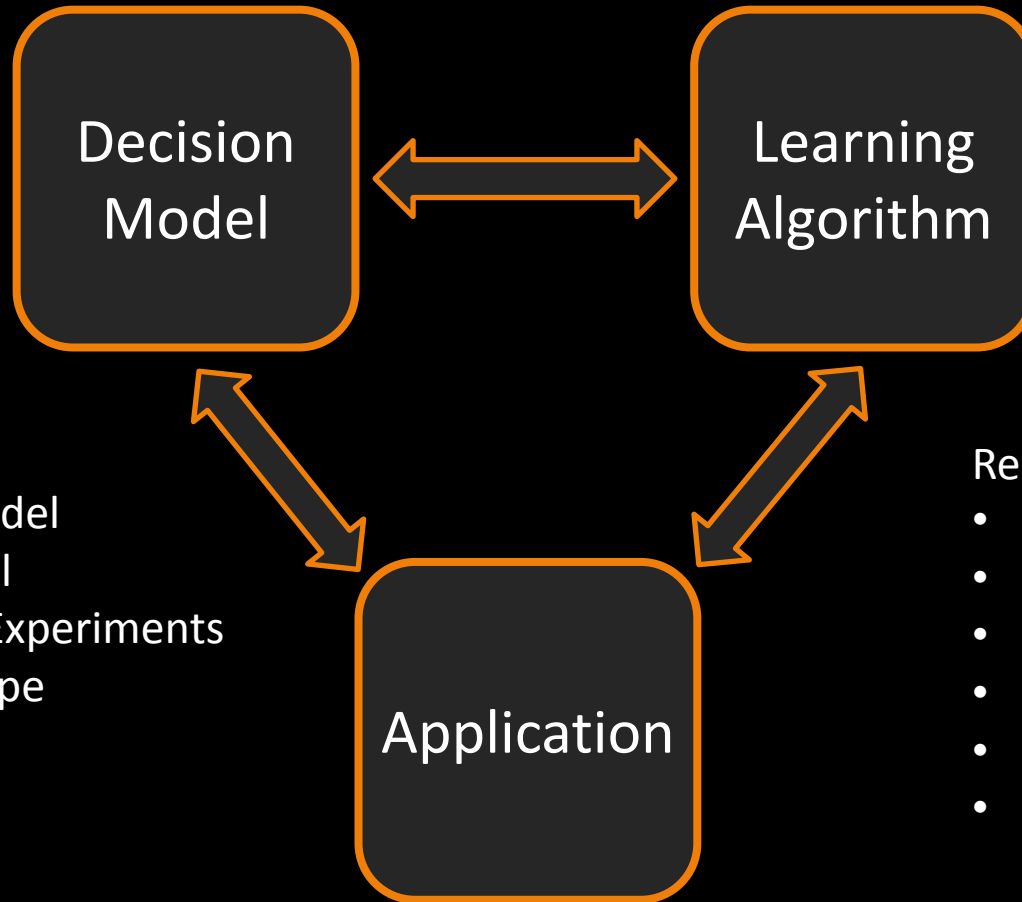


# Summary and Conclusions



- Observed Data  $\neq$  Training Data
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Dueling Bandits
    - $\rightarrow$  Model: Pairwise comparison test  $P(y_i \succ y_j \mid U(y_i) > U(y_j))$
    - $\rightarrow$  Algorithm: Interleaved Filter 2,  $O(|Y| \log(T))$  regret
  - Coactive Learning
    - $\rightarrow$  Model: for given  $y$ , user provides  $\bar{y}$  with  $U(\bar{y} \mid x) > U(y \mid x)$
    - $\rightarrow$  Algorithm: Preference Perceptron,  $O(\|w\| T^{0.5})$  regret Data
  - Propensity Risk Minimization
    - $\rightarrow$  Partial information learning in the Batch setting

# Learning from Human Decisions



## Design Space:

- Decision Model
- Utility Model
- Interaction Experiments
- Feedback Type
- Regret
- Applications

## Related Fields:

- Micro Economics
- Decision Theory
- Econometrics
- Psychology
- Communications
- Cognitive Science

Contact: [tj@cs.cornell.edu](mailto:tj@cs.cornell.edu)

Software + Papers: [www.joachims.org](http://www.joachims.org)