

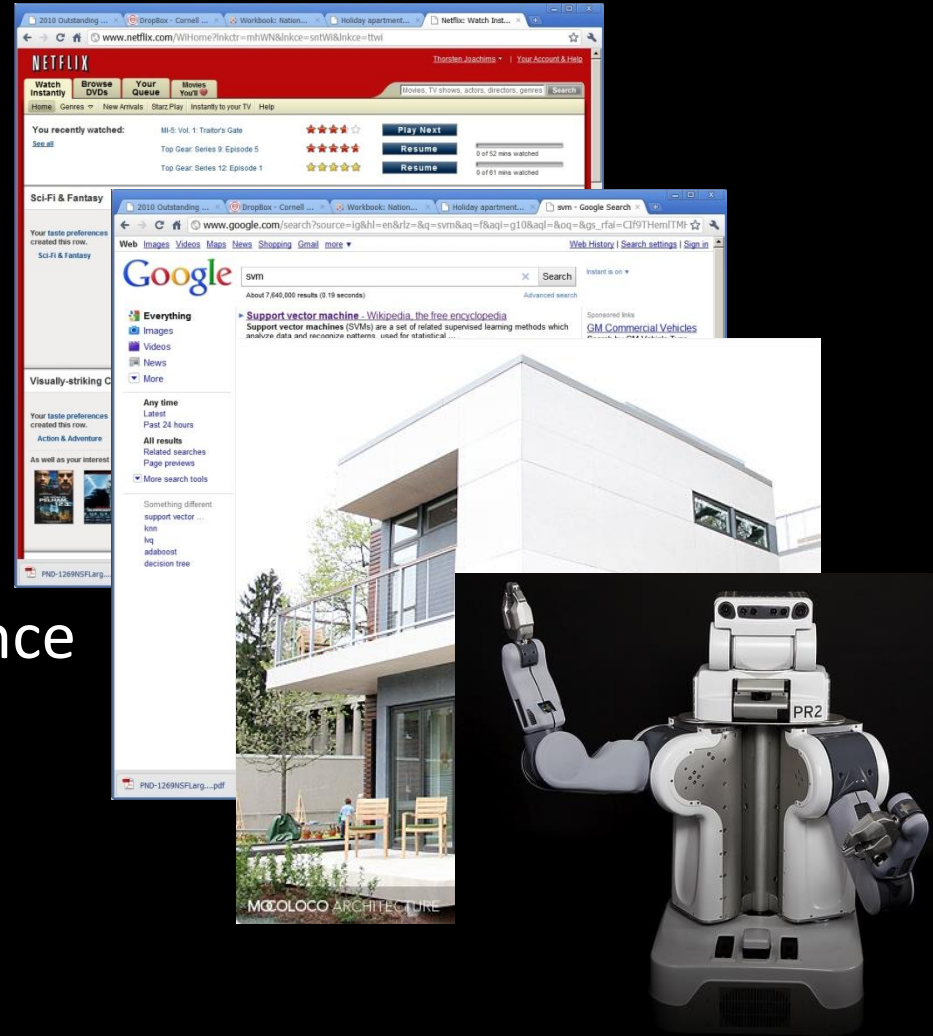
# Online Structured Prediction via Coactive Learning

P. Shivaswamy, T. Joachims

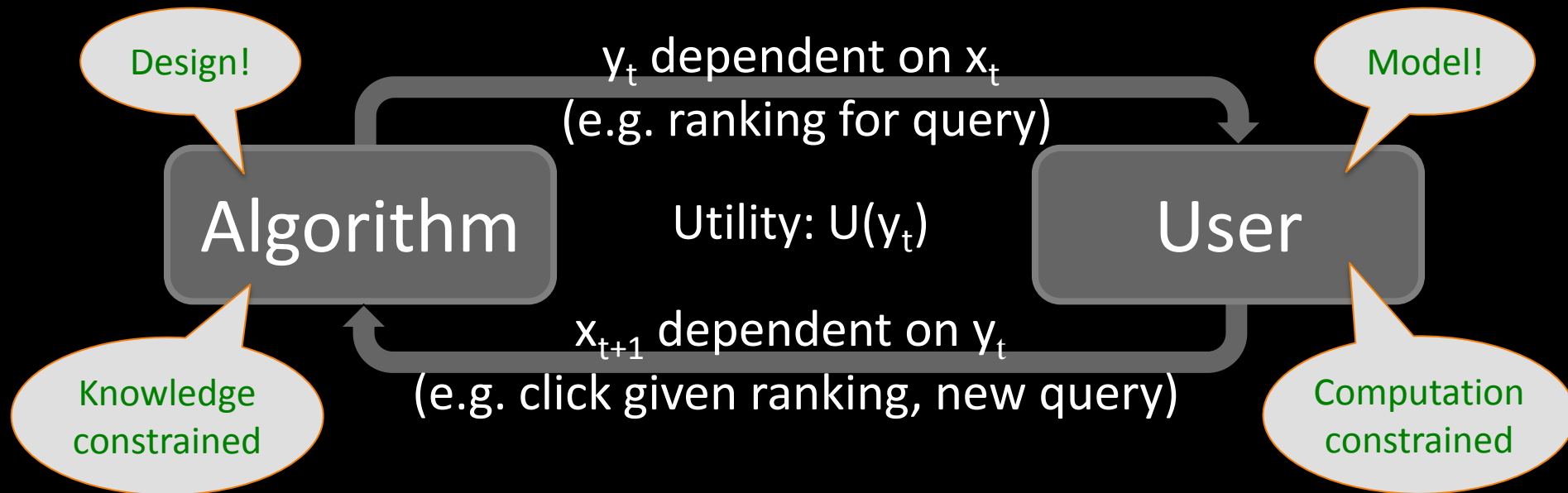
Department of Computer Science  
Cornell University

# User-Facing Machine Learning

- Examples
  - Search Engines
  - Netflix
  - Smart Home
  - Robot Assistant
- Learning
  - Gathering and maintenance of knowledge
  - Measure and optimize performance
  - Personalization

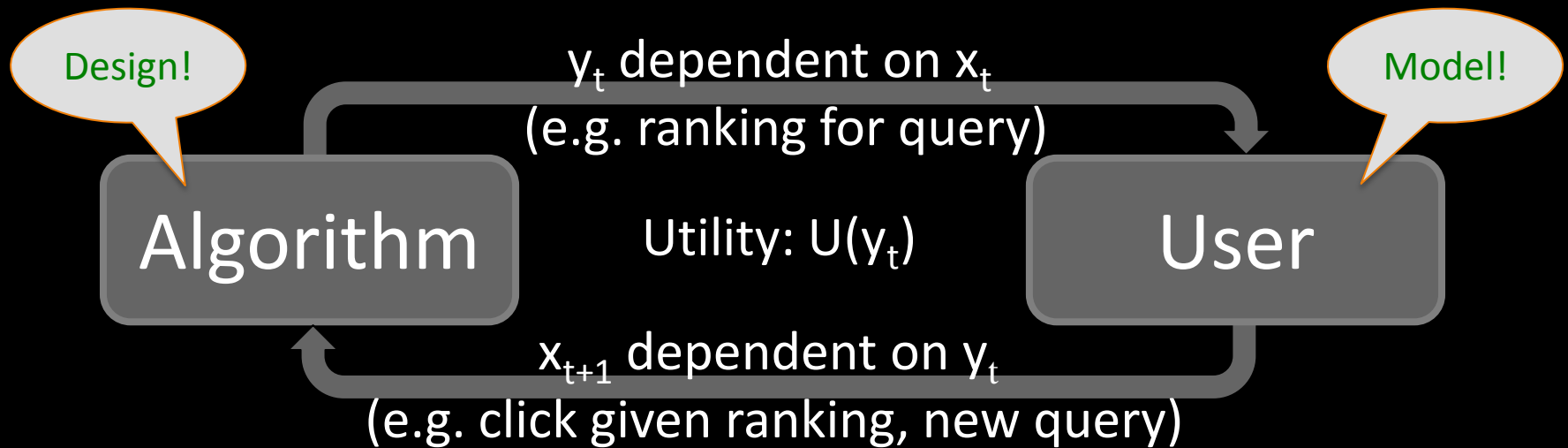


# Interactive Learning System



- Observed Data  $\neq$  Training Data
  - Observed data is user's decisions
  - Need to understand decision process to infer feedback
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm

# Interactive Learning System



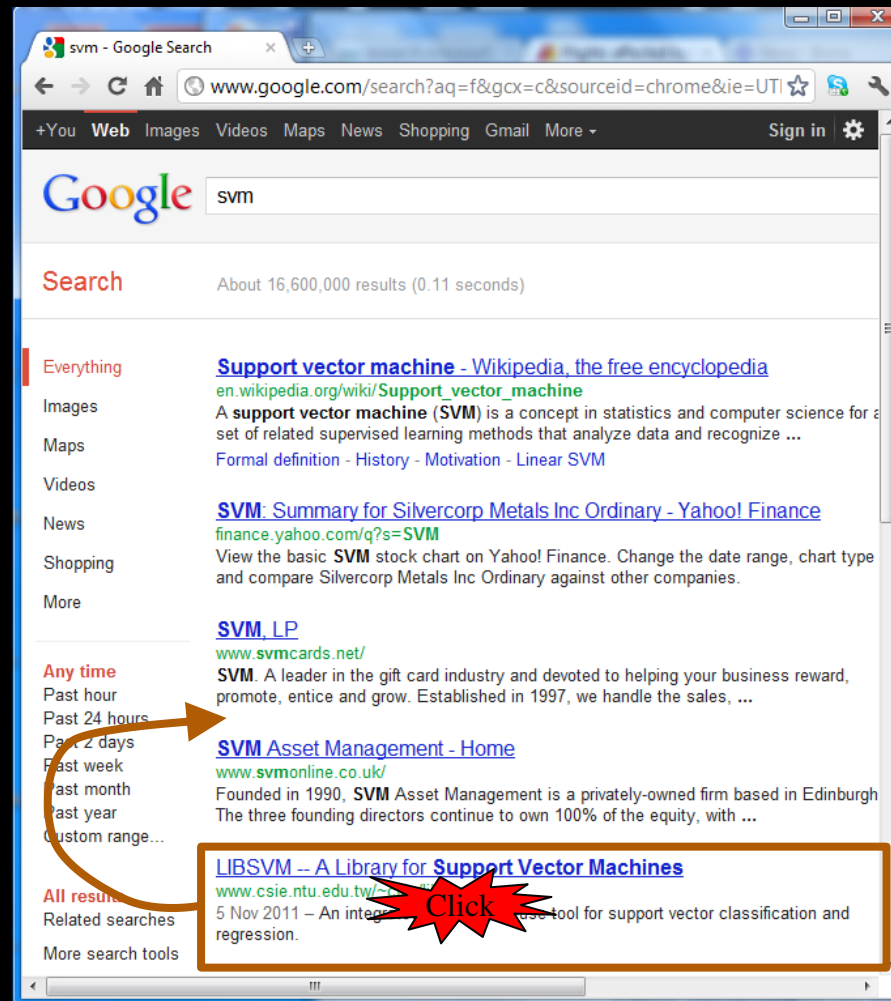
- Observed Data  $\neq$  Training Data ✓
- Decisions  $\rightarrow$  Feedback  $\rightarrow$  Learning Algorithm
  - Model the users decision process to extract feedback ✓
    - $\rightarrow$  Pairwise comparison test  $P(y_i \succ y_j \mid U(y_i) > U(y_j))$
  - Design learning algorithm for this type of feedback ✓
    - $\rightarrow$  Dueling Bandits problem and algorithms (e.g. IF1 and IF2)

# Who does the exploring? Example 1

The image shows a screenshot of the Netflix website interface. The browser address bar indicates the URL `movies.netflix.com/WiMovie/Lie_to_Me/70140406?trkid=13462049`. The page features a red navigation bar with the Netflix logo, '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, a pagination control shows '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. The footer contains copyright information for 2013 Netflix, Inc., and a list of links including Membership, Subtitles & Captions, Test Participation, Gifts, Buy / Redeem, Support, Company, About Us, Affiliates, Investor Relations, Media Center, Jobs, Contact Us, and Blog. A 'Service Code' section at the bottom states: '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



# Who does the exploring? Example 3

The image shows two browser windows side-by-side. The left window displays a Google search for 'svm' with approximately 16,600,000 results. The right window displays 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, specifically highlighting a link to 'SV Meppen 1912 e.V. - Offizielle Webseite-'. A red starburst with the word 'Click' is placed over this link.

**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 analyzes data and recognizes patterns. Formal definition - History - Motivation - Linear SVM

**Images**

**Maps**

**Videos**

**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 data and compare Silvercorp Metals Inc Ordinary against other companies

**Shopping**

**More**

**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

**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.

**Images**

**Maps**

**Videos**  
[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, ...

**News**  
[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

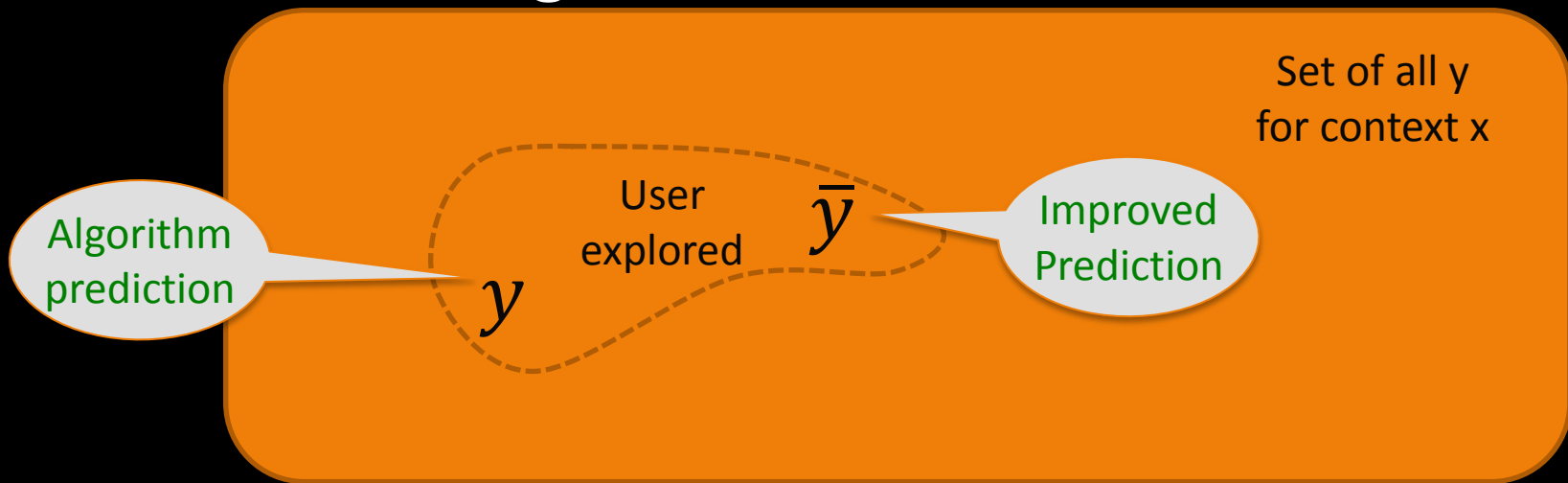
**Shopping**

**More**  
[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~~.

$\bar{y}_t$

# Coactive Learning Model

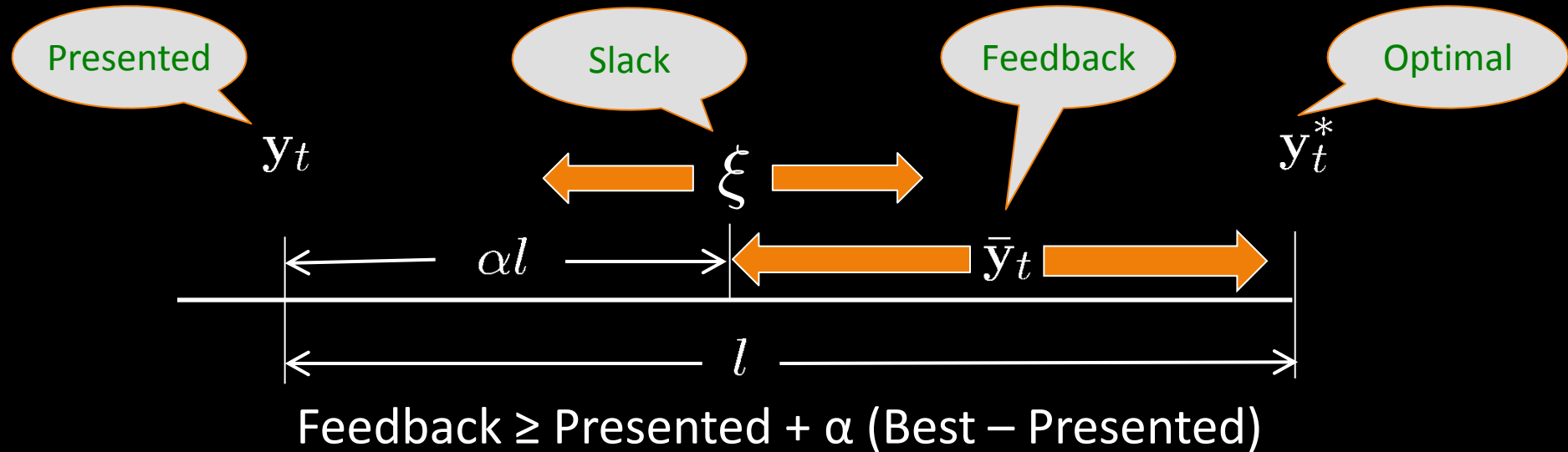
- Unknown Utility Function:  $U(y|x)$ 
    - Boundedly rational user
  - Algorithm/User Interaction:
    - LOOP FOREVER
      - Observe context  $x$  (e.g. query)
      - Learning algorithm presents  $y$  (e.g. ranking)
      - User returns  $\bar{y}$  with  $U(\bar{y}|x) > U(y|x)$
      - $\text{Regret} = \text{Regret} + [ U(y^* | x) - U(y | x) ]$
  - Relationship to other online learning models
    - Expert setting: receive  $U(y|x)$  for all  $y$
    - Bandit setting: receive  $U(y|x)$  only for selected  $y$
    - Dueling bandits: for selected  $y$  and  $\bar{y}$ , receive  $U(\bar{y}|x) > U(y|x)$
    - Coactive setting: for selected  $y$ , receive  $\bar{y}$  with  $U(\bar{y}|x) > U(y|x)$
- Never revealed:
- cardinal feedback
  - optimal  $y^*$
- Loss for prediction  $\hat{y}$
- Optimal prediction  $y^* = \text{argmax}_y \{ U(x, y) \}$

# Preference Perceptron: Algorithm

- Model
  - Linear model of user utility:  $U(y|x) = w^\top \phi(x,y)$
- Algorithm
  - Set  $w_1 = 0$
  - FOR  $t = 1$  TO  $T$  DO
    - Observe  $x_t$
    - Present  $y_t = \operatorname{argmax}_y \{ w_t^\top \phi(x_t, y) \}$
    - Obtain feedback  $\bar{y}_t$
    - 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)

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

# $\alpha$ -Informative Feedback



- **Definition: Strict  $\alpha$ -Informative Feedback**

$$U(\mathbf{x}_t, \bar{y}_t) \geq U(\mathbf{x}_t, y_t) + \alpha(U(\mathbf{x}_t, y_t^*) - U(\mathbf{x}_t, y_t))$$

- **Definition:  $\alpha$ -Informative Feedback**

$$U(\mathbf{x}_t, \bar{y}_t) = U(\mathbf{x}_t, y_t) + \alpha(U(\mathbf{x}_t, y_t^*) - U(\mathbf{x}_t, y_t)) - \xi_t$$

Slacks both  
pos/neg

# Preference Perceptron: Regret Bound

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

- Theorem

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

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

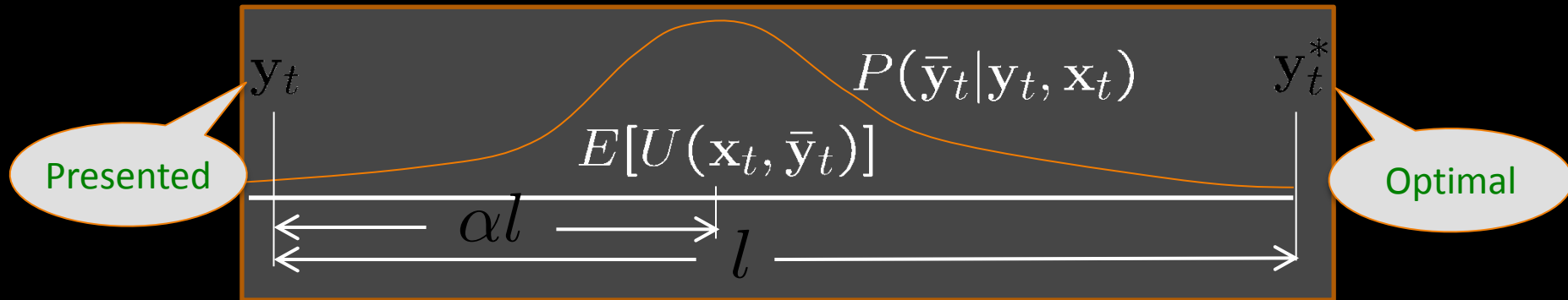
- Other Algorithms and Results

- Feedback that is  $\alpha$ -informative only in expectation
- General convex loss functions of  $U(\mathbf{y}^*|\mathbf{x}) - U(\hat{\mathbf{y}}|\mathbf{x})$
- Regret that scales  $\log(T)/T$  instead of  $T^{-0.5}$  for strongly convex

noise

→ zero

# Expected $\alpha$ -Informative Feedback



- **Definition: Expected  $\alpha$ -Informative Feedback**

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

- **Theorem: Coactive Pref Perceptron achieves**

$$E[REG_T] \leq \frac{1}{\alpha T} \sum_{t=1}^T \bar{\xi}_t + \frac{2R\|w\|}{\alpha\sqrt{T}}$$

# Lower Bound

- Theorem: For any coactive learning algorithm  $A$  with linear utility, there exist  $\mathbf{x}_t$ , objects  $Y$ , and  $\mathbf{w}$  such that  $\text{REG}_T$  of  $A$  in  $T$  steps is  $\Omega(1/T^{0.5})$ .

# 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

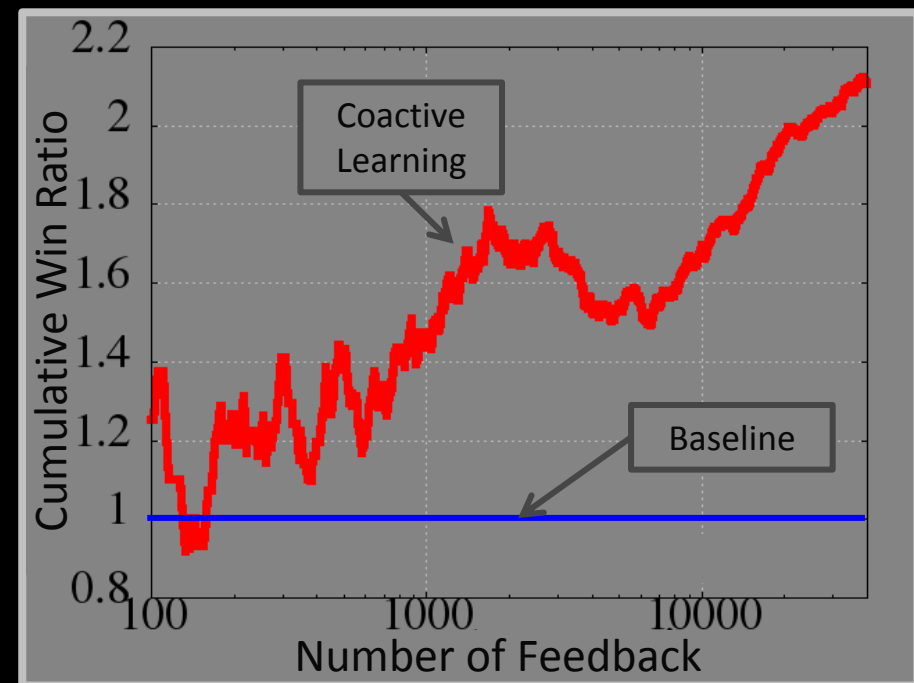
- Construct  $\bar{y}_t$  from  $y_t$  by moving clicked links one position higher.

## Baseline

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

## Evaluation

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





# Related Models

- **Ordinal Regression**

(Crammer & Singer 2001)

- Examples:  $(x_i, r_i)$ ,  
 $r_i$  is numeric rank

- **Pair Preference Learning**

(Herbrich et al., 1999; Freund et al. 2003)

- Examples:  $(x_i, x_i')$
- i.i.d. assumption, batch

- **Ranking**

(Joachims, 2002; Liu 2009)

- Examples:  $(x_i, y_i^*)$ ,  
 $y_i^*$  is optimal ranking
- Structured Prediction, list-wise ranking

- **Expert Model**

- Cardinal feedback for all arms / optimal  $y_i^*$

- **Bandit Model**

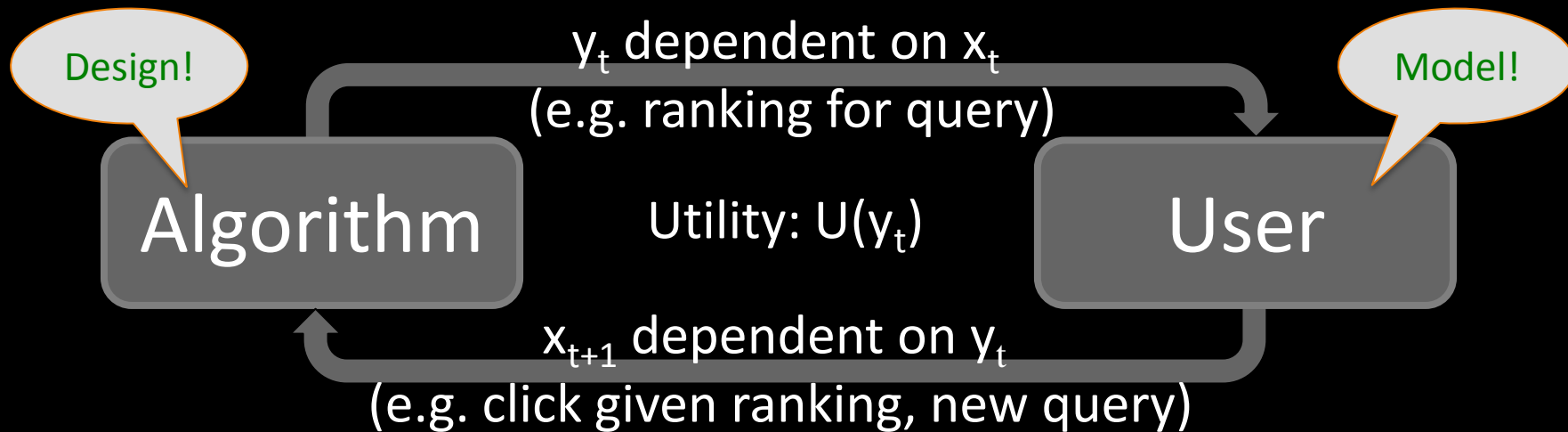
- Cardinal feedback only for chosen arm

- **Dueling Bandit Model**

(Yue et al. 2009; Yue, Joachims 2009)

- Preference feedback between two arms chosen by algorithm

# Summary and Conclusions



	Utility model	Decision model	Actions $y_t$ / Experiment	Feedback	Exploration	Regret
Dueling Bandits	Ordinal	Noisy rational	Comparison pairs	Noisy comparison	Algorithm	Lost comparisons
Coactive Learning	Linear	Bounded rational	Structured object	$\alpha$ -informative $\bar{y}$	User	Cardinal utility
⋮	⋮	⋮	⋮	⋮	⋮	⋮