Online Learning from User Interactions through Interventions

CS 7792 - Fall 2016

Thorsten Joachims

Department of Computer Science & Department of Information Science
Cornell University

Interactive Learning Systems

• Examples
  – Search engines
  – Entertainment media
  – E-commerce
  – Smart homes / robots

• Learning
  – Gathering and maintenance of knowledge
  – Measure and optimize performance
  – Personalization

Interventions
Interactive Learning System

- Information Elicitation from the User
  - Via generative behavioral model
  - Via information-elicitation interventions

- Online Learning with Interventions
  - Dueling Bandits: Algorithm-driven exploration
  - Coactive Learning: User-driven exploration

response $y_t$ dependent on $x_t$
(e.g. ranking for query)

Utility: $U(y_t)$

command $x_t$ and feedback $\delta_t$
(e.g. query, click given ranking)
Decide between two Ranking Functions

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

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/

$U(tj,"SVM",y_1)$

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.freeservers.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/

$U(tj,"SVM",y_2)$
# Measuring Utility

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

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

• None of the absolute metrics reflects expected order.
• Most differences not significant after one month of data.
• Analogous results for Yahoo! Search with much more data [Chapelle et al., 2012].
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]

$\text{argmax } y \in \text{Top}_k U(y)$
Balanced Interleaving

Interleaving($y_1,y_2$)

Interpretation: ($y_1 \succ y_2$) $\iff$ clicks(topk($y_1$)) > clicks(topk($y_2$))

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

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 $k_1=k_2 \pm 1$.

Model of User:
Better retrieval functions is more likely to get more clicks.

1. Kernel Machines
   http://svm.first.gmd.de/
2. Support Vector Machine
   http://jbolivar.freeservers.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/
6. SVM-Light Support Vector Machine
   http://ais.gmd.de/~thorsten/svm_light/
7. Support Vector Machine and Kernel ... References
   http://svm.research.bell-labs.com/SVMrefs.html
8. Lucent Technologies: SVM demo applet
   http://svm.research.bell-labs.com/SVT/SVMsvt.html
9. Royal Holloway Support Vector Machine
   http://svm.dcs.rhbnc.ac.uk
10. SVM-Light Support Vector Machine
    http://ais.gmd.de/~thorsten/svm_light/

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

1. Kernel Machines
   http://svm.first.gmd.de/
2. SVM-Light Support Vector Machine
   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.rhbnc.ac.uk
6. Archives of SUPPORT-VECTOR-MACHINES...
   http://www.jiscmail.ac.uk/lists/SUPPORT...
7. Lucent Technologies: SVM demo applet
   http://svm.research.bell-labs.com/SVT/SVMsvt.html
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

$\Rightarrow$ Ten interleaved queries are equivalent to one manually judged query.

[Radlinski & Craswell, 2010]
Interactive Learning System

- **Information Elicitation from the User**
  - Via generative behavioral model
  - Via information elicitation interventions

- **Online Learning with Interventions**
  - Dueling Bandits: Algorithm-driven exploration
  - Coactive Learning: User-driven exploration
Learning on Operational System

• Example: 4 retrieval functions: \( A > B >> C > D \)
  – 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

[Yue, Broder, Kleinberg, Joachims, 2010]
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**

\[
\text{InterleavedFilter1}(T, W = \{f_1 \ldots f_K\})
\]

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

<table>
<thead>
<tr>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f' = f_3 )</th>
<th>( f_4 )</th>
<th>( f_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>8/2</td>
<td>7/3</td>
<td>4/6</td>
<td>( \times )</td>
<td>( \times )</td>
</tr>
<tr>
<td>13/2</td>
<td>11/4</td>
<td>( \times )</td>
<td>( \times )</td>
<td>XX</td>
</tr>
<tr>
<td>( f' = f_1 )</td>
<td>( f_2 )</td>
<td>( f_4 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0/0</td>
<td>0/0</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

**Related Algorithms:**

[Hofmann, Whiteson, Rijke, 2011] [Yue, Joachims, 2009] [Yue, Joachims, 2011] [Yue et al., 2009]
Assumptions

- Preference Relation: $f_i \succ f_j \iff P(f_i \succ f_j) = 0.5 + \varepsilon_{i,j} > 0.5$
- Weak Stochastic Transitivity: $f_i \succ f_j$ and $f_j \succ f_k \Rightarrow f_i \succ f_k$
- Strong Stochastic Transitivity: $\varepsilon_{i,k} \geq \max\{\varepsilon_{i,j}, \varepsilon_{j,k}\}$
- Stochastic Triangle Inequality: $f_i \succ f_j \succ f_k \Rightarrow \varepsilon_{i,k} \leq \varepsilon_{i,j} + \varepsilon_{j,k}$

Theorem: IF2 incurs expected average regret bounded by

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

- $\varepsilon_{1,2} = 0.01$ 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

response $y_t$ dependent on $x_t$
(e.g. ranking for query)

Utility: $U(y_t)$

command $x_t$ and feedback $\delta_t$
(e.g. query, click given ranking)

• Information Elicitation from the User
  – Via generative behavioral model
  – Via information-elicitation interventions

• Online Learning with Interventions
  – Dueling Bandits: Algorithm-driven exploration
  – Coactive Learning: User-driven exploration
Who does the exploring? Example 1
Who does the exploring?

Example 2
Who does the exploring?

Example 3
Coactive Feedback Model

- Interaction: given $x$
  
- Feedback:
  - Improved prediction $\tilde{y}_t$
    \[ U(\tilde{y}_t | x_t) > U(y_t | x_t) \]
  - Supervised learning: optimal prediction $y_t^*$
    \[ y_t^* = \arg\max_y U(y | x_t) \]
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.
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 $y_t = \arg\max_y \{ w_t^T \phi(x_t,y) \}$
    – Obtain feedback $\tilde{y}_t$ from user
    – Update $w_{t+1} = w_t + \phi(x_t,\tilde{y}_t) - \phi(x_t,y_t)$

• This may look similar to a multi-class Perceptron, but
  – Feedback $\tilde{y}_t$ 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 $y^*$

[Shivaswamy, Joachims, 2012]
Coactive Perceptron: Regret Bound

• Model
  \( U(y|x) = w^T \phi(x, y) \), 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
  For user feedback \( \bar{y} \) that is \( \alpha \)-informative in expectation, the expected average regret of the Preference Perceptron is bounded by
  \[ E \left[ \frac{1}{T} \sum_{t=1}^{T} U(y_t^*|x) - U(y_t|x) \right] \leq \frac{1}{\alpha T} \sum_{t=1}^{T} \xi_t + \frac{2R||w||}{\alpha \sqrt{T}} \]

[Shivaswamy, Joachims, 2012]
Preference Perceptron: Experiment

Experiment:
- Automatically optimize Arxiv.org Fulltext Search

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

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

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

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

[Analogous to DCG]

[Graph showing cumulative win ratio over number of feedback iterations, labeled Coactive Learning and Baseline]
Interactive Learning System

• Information Elicitation Interventions
• 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 $\tilde{y}$ with $U(\tilde{y} \mid x) > U(y \mid x)$
    $\rightarrow$ Algorithm: Preference Perceptron, $O(\|w\| T^{0.5})$ regret
Running Interactive Learning Experiments

1) Build your own system and provide service
   → a lot of work
   → 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
Learning from Human Decisions

- Decision Model
- Learning Algorithm
- Application

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
Software + Papers: www.joachims.org