DYNAMIC SEARCH MODELS AND APPLICATIONS

Jiyun Luo

Department of Computer Science
Georgetown University

Dissertation Committee
Prof. Grace Hui Yang (advisor)
Prof. Ben Carterette
Prof. Nathan Schneider
Prof. Lisa Singh
**Dynamic Search**

E.g. Find what city and state Dulles airport is in, what shuttles ride-sharing vans and taxi cabs connect the airport to other cities, what hotels are close to the airport, what are some cheap off-airport parking, and what are the metro stops close to the Dulles airport.

- $q_1$ – "dulles hotels"
- $q_2$ – "dulles airport"
- $q_3$ – "dulles airport location"
- $q_4$ – "dulles metrostop"

- The information need
  - complex
  - may change and evolve during the search process

- Trial-and-error
  - through repeated, varied attempts until success
  - or is stopped by the user
Another view of Dynamic Search

- Dynamic Search as a path searching in an unknown information space
  - the final goal: may be unclear at the beginning and may evolve
  - a search path: queries, retrieved docs, and examined docs.
  - coworkers -> the user and the search engine
CHARACTERISTICS OF DYNAMIC SEARCH

○ Trial-and-error

- $q_1$ – "dulles hotels"
- $q_2$ – "dulles airport"
- $q_3$ – "dulles airport location"
- $q_4$ – "dulles metrostop"
CHARACTERISTICS OF DYNAMIC SEARCH

- Rich interactions
  - Query formulation
  - Document clicks
  - Document examination
  - eye movement
  - mouse movements
  - etc.
CHARACTERISTICS OF DYNAMIC SEARCH

- Temporal dependency

![Diagram showing the process of dynamic search with iterations and feedback loops.]
REINFORCEMENT LEARNING (RL)

- Fits well in this trial-and-error setting

- It is to learn from repeated, varied attempts which are continued until success.

- The learner (also known as agent) learns from its dynamic interactions with the world
  - rather than from a labeled dataset as in supervised learning.

- It assumes that the system's current state depend on the previous state and action in a non-deterministic manner
PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)

\( s_0 \) \( s_1 \) \( s_2 \) \( s_3 \)  
\( a_0 \) \( o_1 \) \( a_1 \) \( o_2 \) \( a_2 \) \( o_3 \)  
\( r_0 \) \( r_1 \) \( r_2 \)  

\(^1\)R. D. Smallwood et. al., ‘73
PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)

- Elements
  - Hidden states

\[^{1}\text{R. D. Smallwood et. al., ‘73} \]
PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)

- Elements
  - Hidden states
  - Beliefs

---

1 R. D. Smallwood et. al., ‘73
**Partially Observable Markov Decision Process (POMDP)**

- **Elements**
  - Hidden states
  - Beliefs
  - Actions

---

1R. D. Smallwood et. al., '73
Partially Observable Markov Decision Process (POMDP)

- Elements
  - Hidden states
  - Beliefs
  - Actions
  - Rewards

1R. D. Smallwood et. al., ‘73
**Partially Observable Markov Decision Process (POMDP)**

- **Elements**
  - Hidden states
  - Beliefs
  - Actions
  - Rewards
  - Observations

---

1R. D. Smallwood et. al., ‘73
Partially Observable Markov Decision Process (POMDP)

- **Elements**
  - Hidden states
  - Beliefs
  - Actions
  - Rewards
  - Observations

- **Property**
  - Markov

\[ s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow \ldots \]

\[ a_0 \rightarrow o_1 \rightarrow r_0 \]
\[ a_1 \rightarrow o_2 \rightarrow r_1 \]
\[ a_2 \rightarrow o_3 \rightarrow r_2 \]

\(^1\text{R. D. Smallwood et. al., '73}\)
PARTIALLY OBSERVABLE MARKOV DECISION PROCESS (POMDP)

- Elements
  - Hidden states
  - Beliefs
  - Actions
  - Rewards
  - Observations

- Property
  - Markov

- Goal
  - Long Term Optimization

---

1R. D. Smallwood et. al., ‘73
EXISTING WORK ON RL IN IR

○ A decision-theoretic model for query extension and result re-ranking (CIKM’05, X. Shen, B. Tan, and C. Zhai)

○ Contextual bandit to balance exploration and exploitation in web search (ECIR’11, K. Hofmann, S. Whiteson, and M. de Rijke)

○ POMDP in ads recommendation (CIKM’12, Yuan et al.)

○ POMDP in multi-page search result re-ranking (WWW’13, Jin et al.)

○ MDP (Query Change Model) for Session Search (SIGIR’13, Guan et al.)

○ MDP for retrieving streaming TV broadcast content (SIGIR’15, Odijk et al.)

○ Multi-Armed Bandit in Query Expansion (CIKM’16, Li et al.)
CHALLENGES OF RL IN IR

○ Formulation of the Problem
  • What are the states and agents in dynamic search?
  • What are the actions of agents?
  • What are the observations of actions?
  • Etc.

○ How to infer searcher decision making states?

○ Efficiency
  • RL training is computational expensive.
    ◦ how to improve its efficiency?
I provide theoretical models to model dynamic search as a dual-agent stochastic game based on POMDP framework, which are able to infer searcher’s decision making states and provide effective and efficient retrieval performance.
Model 1: POMDP for Dynamic Search

- Model dynamic search as a dual-agent stochastic game
  - two-agent POMDP
PRELIMINARIES: PARTIALLY OBSERVABLE MDP (POMDP)

- A POMDP: a tuple $< S, B, \Omega, A, R, T, O>$:
  - States $S$
  - Belief $B$
  - Observations $\Omega$
  - Actions $A$
  - Reward $R = R(s,a)$
PRELIMINARIES: PARTIALLY OBSERVABLE MDP (POMDP)

A POMDP: a tuple $< S, B, \Omega, A, R, T, O>$:

- States $S$
- Belief $B$
- Observations $\Omega$
- Actions $A$
- Reward $R = R(s,a)$

- $O$ is the observation function which represents a probability distribution for making observation $\omega$ given action $a$ and landing the next state $s'$. 
PRELIMINARIES: PARtially OBSERVABLE MDP (POMDP)

- A POMDP: a tuple $< S, B, \Omega, A, R, T, O>$:
  - States $S$
  - Belief $B$
  - Observations $\Omega$
  - Actions $A$
  - Reward $R = R(s,a)$

- $O$ is the observation function which represents a probability distribution for making observation $\omega$ given action $a$ and landing the next state $s'$.

- **Policy** $\pi : \pi(s) \rightarrow a$. It describes the behaviors of an agent.

- **Value function** Given a policy at time $t$, a value function $V$ calculates the expected long term reward starting from state $s$ inductively.
POMDP Components: States Design

- Related Work
  - MDP in session search [SIGIR’13, Guan et al.]
    - one query -> one state

```
"old US coins"

S1  S2  S3  S4  \ldots  Sn

"collecting old US coins"

"selling old US coins"
```
POMDP COMPONENTS: STATES DESIGN

- Related Work
  - POMDP in re-ranking multi-page search result (WWW’13, Jin et al.)
    - one possible ranking list -> one state
    - # of states: factorial of the number of retrieved documents
WHAT DRIVES SEARCH PROCESS FORWARD?

[TREC’13 session 9]

- E.g. I want to know information about old US coins. How to start collecting old US coins? How to sell old US coins? ...
**States**

- **Two dimensions**
  - “relevant dimension” → whether the user thinks the returned documents are relevant
  - “exploration dimension” → whether the user would like to explore another subtopic.

- **Four States**
  - relevant and exploration
    - e.g. scooter price → scooter stores
  - relevant and exploitation
    - e.g. Hartford visitors → Hartford Connecticut tourism
  - non-relevant and exploitation
    - e.g. Philadelphia NYC travel → Philadelphia NYC train
  - non-relevant and exploration
    - e.g. distance New York Boston → maps.bing.com
**Actions**

- Actions from agents to the environment
  - User actions
    - **Query changes**
      - Added query terms \( +\Delta q_t = q_t \setminus q_{t-1} \)
      - Removed query terms \( -\Delta q_t = q_{t-1} \setminus q_t \)
      - Theme terms \( q_{\text{theme}} = \text{Longest Common Subsequence}(q_t, q_{t-1}) \)
    - **Clicks**
      - SAT clicks
      - Clicks
        - Reading a snippet
        - Reading a document
  - Search engine actions
    - Increasing, decreasing and maintaining term weights
    - Adjusting parameters of different search technologies
**Actions**

- Actions between agents for communication (Messages)
  - Messages from users ($\Sigma_u$): the clicked documents ($D_{clicked}$)
  - Messages from search engine ($\Sigma_{se}$): the top k returned documents
**Observation Function**

- The Observation function $O(s_j, a_t, \omega_t)$, defined as $P(\omega_t | s_j, a_t)$, is the probability of observing $\omega_t \in \Omega$ when agents take action $a_t$ and land on state $s_j$.

- **Relevant or Non-relevant**
  - $O(s_t=Relevant, \Sigma_u, \omega_t=Relevant)$
    - given user action $\Sigma_u$ and the current state is “Relevant”, the probability of observing “Relevant” (observing a SAT Click)
**Observation Function**

- The Observation function $O(s_j, a_t, \omega_t)$, defined as $P(\omega_t | s_j, a_t)$, is the probability of observing $\omega_t \in \Omega$ when agents take action $a_t$ and land on state $s_j$.

- Relevant or Non-relevant
  - $O(s_t=\text{Relevant}, \Sigma_u, \omega_t=\text{Relevant})$
    - given user action $\Sigma_u$ and the current state is “Relevant”, the probability of observing “Relevant” (observing a SAT Click)

  \[ \propto P(s_t=\text{Relevant} | \omega_t=\text{Relevant}) P(\omega_t=\text{Relevant} | \Sigma_u) \]

  \[ P(s_t=\text{Relevant} | \omega_t=\text{Relevant}) = \frac{\# \text{ of observed true relevant}}{\# \text{ of observed relevant}} \]
**Observation Function**

- The Observation function \( O(s_j, a_t, \omega_t) \), defined as \( P(\omega_t | s_j, a_t) \), is the probability of observing \( \omega_t \in \Omega \) when agents take action \( a_t \) and land on state \( s_j \).

- **Relevant or Non-relevant**
  - \( O(s_t=\text{Relevant}, \Sigma_u, \omega_t=\text{Relevant}) \)
    - given user action \( \Sigma_u \) and the current state is “Relevant”, the probability of observing “Relevant” (observing a SAT Click)

  \[ \propto P(s_t=\text{Relevant} | \omega_t=\text{Relevant})P(\omega_t=\text{Relevant} | \Sigma_u) \]

  \[ P(s_t=\text{Relevant} | \omega_t=\text{Relevant}) = \frac{\# \text{ of observed true relevant}}{\# \text{ of observed relevant}} \]

  - \( s_t \) is likely to be \[
    \begin{cases} 
      \text{Relevant} & \text{if a SAT Click in } D_{t-1} \\
      \text{Non-Relevant} & \text{otherwise.}
    \end{cases}
  \]
The Observation function

- Exploration or Exploitation
  - $O(s_t = \text{Exploration}, a_u = \Delta q_t, \Sigma_{se} = D_{t-1}, \omega_t = \text{Exploration})$
  - $\propto P(s_t = \text{Exploration}|\omega_t = \text{Exploration})$
  - $\times P(\omega_t = \text{Exploration}|\Delta q_t, D_{t-1})$

- $s_t$ is likely to be
  - Exploration if (added term is from nowhere)
  - or (no added term, but removed term from $q_{t-1}$)
  - Exploitation if (newly added term is from $D_{t-1}$)
  - or (no term change from $q_{t-1}$)
Observation Function

- The Observation function
  - Exploration or Exploitation
    - $O(s_t = \text{Exploration}, a_u = \Delta q_t, \Sigma_{se} = D_{t-1}, \omega_t = \text{Exploration})$
    $$\propto P(s_t = \text{Exploration}|\omega_t = \text{Exploration})$$
    $$\times P(\omega_t = \text{Exploration}|\Delta q_t, D_{t-1})$$
  - $s_t$ is likely to be
    \[
    \begin{cases}
    \text{Exploration} & \text{if (added term is from nowhere)} \\
    \text{or (no added term, but removed term from } q_{t-1}) \\
    \text{Exploitation} & \text{if (newly added term is from } D_{t-1}) \\
    \text{or (no term change from } q_{t-1})
    \end{cases}
    \]

- The final observation function
  - the product of the two observation functions for relevant and exploration
    - assume relevant and exploration are independent variables
**Belief Updates**

- At every search iteration the belief state $b$ is updated when a new observation is obtained.
  - based on the observation function

The belief update function is

$$b_{t+1}(s_j) = P(s_j|\omega_t, a_t, b_t)$$

$$= \frac{O(s_j, a_t, \omega_t) \sum_{s_i \in \mathcal{S}} P(s_j|s_i, a_t, b_t) b_t(s_i)}{P(\omega_t|a_t, b_t)}$$
REWARDS

- The immediate reward
  - clicks and relevance score.
    - it measures the gain received by the user from examining $D_{t-1}$
Rewards

- The immediate reward
  - clicks, relevance score, etc.
    - it measures the gain received by the user from examining $D_{t-1}$

- The long term reward function for the search engine agent

\[
Q_{se}(b, a) = \sum_{s \in S} b(s) R(s, a) + \gamma \sum_{\omega \in \Omega} P(\omega|b, a_{\omega}, \Sigma_{se}) P(\omega|b, \Sigma_{u}) \max_{a} Q_{se}(b', a)
\]

Immediate reward  Discounted future reward
EXPERIMENTS

- Data: TREC 2012 and 2013 Session Tracks
- Corpus: ClueWeb09, ClueWeb12

<table>
<thead>
<tr>
<th>Table 3: TREC Session Dataset Statistics.</th>
<th>TREC’12</th>
<th>TREC’13</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Sessions</td>
<td>98</td>
<td>87</td>
</tr>
<tr>
<td>#Queries</td>
<td>297</td>
<td>442</td>
</tr>
<tr>
<td>Avg. session length</td>
<td>3.03</td>
<td>5.08</td>
</tr>
<tr>
<td>Max session length</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>#Sessions w/ length &lt;4</td>
<td>75</td>
<td>38</td>
</tr>
<tr>
<td>#Sessions w/ length 4~10</td>
<td>22</td>
<td>42</td>
</tr>
<tr>
<td>#Sessions w/ length &gt;10</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>
EXPERIMENTS

- TREC Session Track Task

```xml
<session num="1" starttime="0">
  <topic>
    <interaction num="1" starttime="10.280644">
      <query>wikipedia cosmetic laser treatment</query>
      <results>
        <result rank="1">
          <url>http://www.veindirectory.org/content/varicose_veins.asp</url>
          <title>Varicose Veins - Vein Treatment, Removal, Surgery Information</title>
          <snippet>... more severe problems such as leg pain, ...</snippet>
        </result>
        ...
      </results>
      <clicked>
        <click num="1" starttime="95.603468" endtime="120.565420">
          <rank>10</rank>
        </click>
        ...
      </clicked>
    </interaction>
    <currentquery starttime="252.659006">
      <query>uses for cosmetic laser treatment</query>
    </currentquery>
  </topic>
</session>
```
EXPERIMENTS

- TREC Session Track Task

```xml
<session num="1" starttime="0">
  <topic>
    <interaction num="1" starttime="10.280644">
      <query>wikipedia cosmetic laser treatment</query>
      <results>
        <result rank="1">
          <url>http://www.veindirectory.org/content/varicose_veins.asp</url>
          <title>Varicose Veins - Vein Treatment, Removal, Surgery Information</title>
          <snippet> ... more severe problems such as leg pain, ... </snippet>
        </result>
        ...
      </results>
      <clicked>
        <click num="1" starttime="95.603468" endtime="120.565420">
          <rank>10</rank>
        </click>
        ...
      </clicked>
    </interaction>
    <currentquery starttime="252.659006">
      <query>uses for cosmetic laser treatment</query>
    </currentquery>
  </topic>
</session>
```

- The task is to retrieve a ranked list of 2,000 documents for the last query in a session.
**Baselines**

- Lemur (language model + Dirichlet smoothing)
- PRF (pseudo relevance feedback with top 20 ranked docs)
- Rocchio (relevance feedback using top 10 ranked docs)
- Rocchio-CLK (relevance feedback using previously clicked docs)
- Rocchio-SAT (relevance feedback using previously SAT-clicked docs)
- QCM+DUP [Guan et al. SIGIR’ 13]
- QCM+SAT [Zhang et al. SIGIR’ 13]
- TREC median and best (scores across all runs)

- the original run
  - the retrieval list provide by TREC log data.

- Win-Win (proposed algorithm)
**Metrics**

- **Search Accuracy scores**
  - Normalized Discounted Cumulative Gain (nDCG)
  - nDCG@10
  - Mean Average Precision (MAP)
  - normalized Expected Reciprocal Rank (nERR@10)

- **Immediate Search Accuracy**
  - nDCG@10 and nERR@10 at each search iteration i.
**Search Accuracy**

- Search accuracy on TREC 2012 and 2013 Session

<table>
<thead>
<tr>
<th>Approach</th>
<th>nDCG@10</th>
<th>nDCG</th>
<th>MAP</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemur</td>
<td>0.2474</td>
<td>0.2627</td>
<td>0.1274</td>
<td>0.2857</td>
</tr>
<tr>
<td>TREC median</td>
<td>0.2608</td>
<td>0.2468</td>
<td>0.1440</td>
<td>0.2626</td>
</tr>
<tr>
<td>TREC best</td>
<td>0.3221</td>
<td>0.2865</td>
<td>0.1559</td>
<td>0.3595</td>
</tr>
<tr>
<td>PRF</td>
<td>0.2074</td>
<td>0.2335</td>
<td>0.1065</td>
<td>0.2415</td>
</tr>
<tr>
<td>Rocchio</td>
<td>0.2446</td>
<td>0.2714</td>
<td>0.1281</td>
<td>0.2950</td>
</tr>
<tr>
<td>Rocchio-CLK</td>
<td>0.2916↑</td>
<td>0.2866</td>
<td>0.1449</td>
<td>0.3366</td>
</tr>
<tr>
<td>Rocchio-SAT</td>
<td>0.2889</td>
<td>0.2836</td>
<td>0.1467</td>
<td>0.3254</td>
</tr>
<tr>
<td>QCM+DUP</td>
<td>0.2742</td>
<td>0.2560</td>
<td>0.1537†</td>
<td>0.3221</td>
</tr>
<tr>
<td>QCM SAT</td>
<td>0.3550*†</td>
<td>0.3054</td>
<td>0.1534†</td>
<td>0.1534</td>
</tr>
<tr>
<td>Win-Win</td>
<td>0.2941†</td>
<td>0.2691</td>
<td>0.1346</td>
<td>0.3403</td>
</tr>
</tbody>
</table>

**TREC 2012 Session Track**

- In TREC 2012, win-win search is better than most systems except QCM SAT in terms of nDCG@10.

<table>
<thead>
<tr>
<th>Approach</th>
<th>nDCG@10</th>
<th>nDCG</th>
<th>MAP</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemur</td>
<td>0.1147</td>
<td>0.1758</td>
<td>0.0926</td>
<td>0.1314</td>
</tr>
<tr>
<td>TREC median</td>
<td>0.1531</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>TREC best</td>
<td>0.1952</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PRF</td>
<td>0.1061</td>
<td>0.1701</td>
<td>0.0787</td>
<td>0.1245</td>
</tr>
<tr>
<td>Rocchio</td>
<td>0.1320</td>
<td>0.1924</td>
<td>0.1060</td>
<td>0.1549</td>
</tr>
<tr>
<td>Rocchio-CLK</td>
<td>0.1315</td>
<td>0.1929</td>
<td>0.1060</td>
<td>0.1546</td>
</tr>
<tr>
<td>Rocchio-SAT</td>
<td>0.1147</td>
<td>0.1758</td>
<td>0.0926</td>
<td>0.1314</td>
</tr>
<tr>
<td>QCM+DUP</td>
<td>0.1316</td>
<td>0.1929</td>
<td>0.1060</td>
<td>0.1547</td>
</tr>
<tr>
<td>QCM SAT</td>
<td>0.1186</td>
<td>0.1754</td>
<td>0.0939</td>
<td>0.1425</td>
</tr>
<tr>
<td>Win-Win</td>
<td>0.2026*†</td>
<td>0.2609*†</td>
<td>0.1290*†</td>
<td>0.2328*†</td>
</tr>
</tbody>
</table>
Search accuracy on TREC 2012 and 2013 Session

**TREC 2012 Session Track**

- In TREC 2012, win-win search is better than most systems except QCM SAT in terms of nDCG@10.
- In TREC 2013, win-win is the best run among all systems in terms of all evaluation metrics.

**TREC 2013 Session Track**

- Search accuracy on TREC 2012 and 2013 Session
In TREC 2012, win-win search
- immediate nDCG@10 score > the original run from Iteration 2.
Immediate Search Accuracy

- In TREC 2012, win-win search
  - immediate nDCG@10 score > the original run from Iteration 2.
  - immediate nERR@10 score > the original run from Iteration 3.
In TREC 2012, win-win search
  ○ immediate nDCG@10 score > the original run from Iteration 2.
  ○ immediate nERR@10 score > the original run from Iteration 3.
In TREC 2013, win-win search > the original run in every iteration.
Win-win search's immediate search accuracy increases while the number of search iterations increases
OUTLINE

- A POMDP framework for modeling dynamic search

- **Detecting user struggle state in search**

- A direct policy learning framework

- Implementation of a dynamic search tool
DECISION MAKING STATES IN POMDP

dimension 1
- Relevant
- Non-Relevant

dimension 2
- Exploration
- Exploitation

dimension 3
- Struggle
- Non-Struggle
Without considering struggles, it seems that the user wants to exploit “lunch menu for women”.
Detect User Search Struggle State

21:46:59 Q2: lunch menu for women

• Without considering struggles, it seems that the user wants to exploit “lunch menu for women”.

3/12/15 17:41:49 Q3: light lunch menu ideas
17:42:18 Click: www.southernliving.com/food/enter...

• Actually search for “light lunch menu”
Detect User Search Struggle State

21:46:59 Q2: lunch menu for women

- Without considering struggles, it seems that the user wants to exploit “lunch menu for women”.

3/12/15 17:41:49 Q3: light lunch menu ideas
17:42:18 Click: www.southernliving.com/food/enter...

- Actually search for “light lunch menu”

- Detect struggles can help
  - correctly adjust search strategy

- Challenge: no simple heuristic rules can apply
Solutions

- Applying supervised machine learning to predict user struggle state
  - using user behavioral signals extracted from Yahoo! search logs
  - binary classification of struggling
    - 1 -> Struggle
    - 0 -> Non-struggle
PROBLEM DEFINITION

- **Struggling Session**: a session where a user experiences difficulty in finding information at some moment during the search.

- **Non-Struggle Session**: a user does not experience any struggle at any moment throughout the search session.
  - makes continuous progress towards task completion.
## Examples

<table>
<thead>
<tr>
<th>Session 1</th>
<th>3/9/15 22:11:32</th>
<th>Q1: movies ten Lynchburg VA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22:17:23</td>
<td>Q2: regal movies Lynchburg VA</td>
</tr>
<tr>
<td></td>
<td>22:19:14</td>
<td>Q3: movies ten Lynchburg VA</td>
</tr>
<tr>
<td></td>
<td>22:19:21</td>
<td>Click: <a href="http://www.mrmovietimes.com/movie-the...">www.mrmovietimes.com/movie-the...</a></td>
</tr>
</tbody>
</table>

(a) non-struggle search session

<table>
<thead>
<tr>
<th>Session 2</th>
<th>3/11/15 21:46:21</th>
<th>Q1: lunch menu ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21:46:59</td>
<td>Q2: lunch menu for women</td>
</tr>
<tr>
<td></td>
<td>3/12/15 17:41:49</td>
<td>Q3: light lunch menu ideas</td>
</tr>
<tr>
<td></td>
<td>17:42:18</td>
<td>Click: <a href="http://www.southernliving.com/food/enter...">www.southernliving.com/food/enter...</a></td>
</tr>
</tbody>
</table>

(b) struggling search session

- Session 1: the searcher was smoothly exploring the movie theaters near Lynchburg, VA.
- Session 2: the searcher struggled to formulate a good query at the beginning and it took her another day to finish the task.
DATA ANNOTATION

1 2015-03-13 3:14:17
Query overactive bladder vs interstitial cystitis
Source: strugglev3(0)

2 2015-03-13 3:14:25
Query overactive bladder vs interstitial cystitis
Source: strugglev3(1)

3 2015-03-13 3:14:32
Click www.ic-hope.com/overactive-bladder-or-interstitial-cystitis.html
Source: strugglev3(2)

4 2015-03-13 3:17:07
Click www.medhelp.org/posts/Urogynecology/Interstitial-cystitis-vs-overactive-bladder/show/1204950
Source: strugglev3(3)

5 2015-03-13 3:20:04
Click www.ichelp.org/Page.aspx?pid=918
Source: strugglev3(4)

Search Sessions conducted using mobile devices

- #Sessions: 1,226
- #Struggle: 483
- #Non-struggle: 743

Search Sessions using PCs

- #Sessions: 365
- #Struggle: 170
- #Non-struggle: 195

Does the session indicate the user is struggling?
- Strong YES
- Weak YES
- Uncertain
- Weak NO
- Strong NO

More than one goal in this session

Comment:
### Features

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query Features</strong></td>
<td></td>
<td><strong>Click Diversity Features</strong></td>
<td></td>
</tr>
<tr>
<td>NumQuery</td>
<td># of queries in the current task</td>
<td>UniClickRatio</td>
<td>percentage of unique URLs</td>
</tr>
<tr>
<td>NumUniQuery*</td>
<td># of unique queries</td>
<td>NumRevisitClick*</td>
<td># of revisited Clicks</td>
</tr>
<tr>
<td>AvgTerm</td>
<td>avg. # of terms per query</td>
<td>RevisitClickRatio*</td>
<td>percentage of revisited Clicks</td>
</tr>
<tr>
<td>AvgChar</td>
<td>avg. # of characters per query</td>
<td>UniDomainRatio</td>
<td>percentage of clicked unique domains</td>
</tr>
<tr>
<td>ManualQueryRatio</td>
<td>ratio of manually typed queries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SuggestQueryRatio</td>
<td>ratio of suggested queries (automatically corrected, suggested, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Query Transition Features</strong></td>
<td></td>
<td><strong>Time Features</strong></td>
<td></td>
</tr>
<tr>
<td>AvgQ1Sim</td>
<td>avg. similarity between all queries and the first query</td>
<td>AvgCDwellTime*</td>
<td>log(1 + avg. dwell time per click include and exclude clicks for the last query)</td>
</tr>
<tr>
<td>AvgAddTerm</td>
<td>avg. # of terms added to next query</td>
<td>TotalClickDwell</td>
<td>total dwell time of all clicks</td>
</tr>
<tr>
<td>MaxQueryPos*</td>
<td>the longest query’s position</td>
<td>AvgQDwellTime*</td>
<td>log(1 + avg. dwell time per query include and exclude the last query)</td>
</tr>
<tr>
<td>NumQSpecify</td>
<td># of queries where one or more terms are deleted from previous query</td>
<td>DurationSAT1*</td>
<td>log(1 + time passed until 1st SAT-click)</td>
</tr>
<tr>
<td>NumQGeneral</td>
<td># of queries where one or more terms are removed from previous query</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StdQueryLength*</td>
<td>standard deviation of query length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgMatchTerm</td>
<td>avg. # of terms that exactly match the previous query</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgDelTerm</td>
<td>avg. # of deleted query terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgQueryEdit*</td>
<td>avg. edit distance of neighbor queries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DiffQ1Qavg*</td>
<td>difference between the first query length and the avg. query length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgQuerySim</td>
<td>avg. query similarity per query pair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgSubTerm</td>
<td>avg. # of substituted terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Click Features</strong></td>
<td></td>
<td><strong>Query &amp; Click History Features</strong></td>
<td></td>
</tr>
<tr>
<td>NumSCClick*</td>
<td>total &amp; avg. # of clicked search cards</td>
<td>AvgQClickFreq*</td>
<td>log(1 + avg. click frequency per clicked URL under the same query)</td>
</tr>
<tr>
<td>NumClick</td>
<td>total and avg. # of clicks per query</td>
<td>AvgQFastBackCTR</td>
<td>avg. fast-back click (dwell time &lt; 15s) through rate per query</td>
</tr>
<tr>
<td>NumSATClick*</td>
<td>total &amp; avg. # of SAT-clicks per query</td>
<td>AvgQClickEntropy</td>
<td>avg. entropy of query click distribution.</td>
</tr>
<tr>
<td>NumPagination*</td>
<td>total and avg. times of pagination</td>
<td>AvgQClickFreq*</td>
<td>log(1 + avg. click frequency per clicked URL in the current task)</td>
</tr>
<tr>
<td>isClickEnd*</td>
<td>whether the task ends with a click</td>
<td>AvgQueryCTR</td>
<td>avg. clickthrough rate per query</td>
</tr>
<tr>
<td>AbandonRatio</td>
<td>percentage of abandoned queries</td>
<td>AvgQSatCTR</td>
<td>avg. SAT clickthrough rate per query</td>
</tr>
<tr>
<td>NumEvent*</td>
<td># of events (click, pagination, etc.)</td>
<td>AvgQueryFreq</td>
<td>log(1 + avg. query occurrences in the external log of search using PCs)</td>
</tr>
<tr>
<td>NumAdsClick</td>
<td>total and avg. ads clicked per query</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NoneClickSeq*</td>
<td>max &amp; avg. # of sequential queries without clicks</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topic Features</strong></td>
<td></td>
<td><strong>Topic Features</strong></td>
<td></td>
</tr>
<tr>
<td>TaskTopic</td>
<td>task’s major ODP topic label</td>
<td>TopicRichness</td>
<td>total number of unique topics in a task</td>
</tr>
<tr>
<td>TopicEntropy</td>
<td></td>
<td>TopicEntropy</td>
<td>entropy of topic distribution</td>
</tr>
<tr>
<td><strong>Mobile Features</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MobileModel*</td>
<td>mobile model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MobileBrand*</td>
<td>mobile brand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MobileOS*</td>
<td>mobile operating system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ScreenSize*</td>
<td>mobile screen size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browser*</td>
<td>mobile browser type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NumSCImpression*</td>
<td>avg. # of search cards showed per query</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NumScreenBtmAct*</td>
<td>total and avg. num of actions triggered at the bottom area of a phone screen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* new features, otherwise features used in baseline [Hassan et al. WSDM’ 14]
EXPERIMENTS

- **ZeroRule**: label all instances using the majority label of the datasets
- **Baseline**: MART model trained on features in [Hassan et al. WSDM’ 14]
- SVM
- LR
- MART
  - 10-fold Cross-validation for all models

**Metrics**
- Classification Accuracy
- Precision and Recall for the positive label (“Struggle”)
- Precision and Recall for the negative label (“Grounded”)
### Classification Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>accuracy</th>
<th>pos. precision</th>
<th>pos. recall</th>
<th>neg. precision</th>
<th>neg. recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZeroRule</td>
<td>0.6060</td>
<td>–</td>
<td>0.0000</td>
<td>0.6060</td>
<td>1.0000</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.7170</td>
<td>0.6699</td>
<td>0.5810</td>
<td>0.7442</td>
<td>0.8180</td>
</tr>
<tr>
<td>SVM</td>
<td>0.7341†</td>
<td>0.7123‡</td>
<td>0.5610</td>
<td>0.7449</td>
<td>0.8562‡</td>
</tr>
<tr>
<td>LR</td>
<td>0.7402‡</td>
<td>0.6896</td>
<td>0.6372‡</td>
<td>0.7702†</td>
<td>0.8206</td>
</tr>
<tr>
<td>MART</td>
<td>0.7610‡</td>
<td>0.7263‡</td>
<td>0.6485‡</td>
<td>0.7794†</td>
<td>0.8444‡</td>
</tr>
</tbody>
</table>

- Among above models, MART is the most effective training model in mobile search.
## Classification Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>pos. precision</th>
<th>pos. recall</th>
<th>neg. precision</th>
<th>neg. recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZeroRule</td>
<td>0.6060</td>
<td>-</td>
<td>0.0000</td>
<td>0.6060</td>
<td>1.0000</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.7170</td>
<td>0.6699</td>
<td>0.5810</td>
<td>0.7442</td>
<td>0.8180</td>
</tr>
<tr>
<td>SVM</td>
<td>0.7341†</td>
<td>0.7123‡</td>
<td>0.5610</td>
<td>0.7449</td>
<td>0.8562‡</td>
</tr>
<tr>
<td>LR</td>
<td>0.7402‡</td>
<td>0.6896</td>
<td>0.6372‡</td>
<td>0.7702†</td>
<td>0.8206</td>
</tr>
<tr>
<td>MART</td>
<td>0.7610‡</td>
<td>0.7263‡</td>
<td>0.6485‡</td>
<td>0.7794†</td>
<td>0.8444‡</td>
</tr>
</tbody>
</table>

- Among above models, MART is the most effective training model in mobile search.
- It is consistent for search using PC platforms.

| PC—MART    | 0.8131   | 0.7632         | 0.8790      | 0.8710         | 0.7718      |
OUTLINE

- A POMDP framework for modeling dynamic search
- Detecting user struggle state in search
- A direct policy learning framework
- Implementation of a dynamic search tool
MODEL 2: DIRECT POLICY LEARNING

<table>
<thead>
<tr>
<th>Approach</th>
<th>Wall Clock</th>
<th>CPU Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>lemur</td>
<td>$2.6 \times 10^3 s$</td>
<td>$7.0 \times 10^{12}$</td>
</tr>
<tr>
<td>qcm [Guan et al. SIGIR’ 13]</td>
<td>$5.6 \times 10^3 s$</td>
<td>$1.5 \times 10^{13}$</td>
</tr>
<tr>
<td>winwin</td>
<td>$2.5 \times 10^4 s$</td>
<td>$6.8 \times 10^{13}$</td>
</tr>
</tbody>
</table>

- Addresses high complexity of RL in IR
  - directly learns mappings from observations to actions
    - skips states, beliefs
  - flatten the model structure (a more down-to-the-earth model)
    - ... but, still complex enough to be interesting
  - less model complexity leads to higher efficiency
A DIRECT POLICY LEARNING FRAMEWORK

- At each search iteration, the search engine maximizes long-term rewards (value function)

\[ V_{\theta}(s_0) = E\left( \sum_{t=0}^{\infty} \gamma^t r(t) | s_0 \right) \]

- Learns a direct mapping from observations to actions by gradient descent

\[ V_{\theta}(s_0) = \sum_{t=0}^{\infty} \gamma^t \sum_{h \in H} P(h|\theta) r(t, h) \]
Defining a History

- History: the record of a session from the search iteration 0 to the current iteration \( t \)

- A chain of events happening in a session

  - the dynamic changes of states, actions, observations, and rewards in a session

\[
\begin{align*}
    h_t &= [h_{t-1}, C_t, T_t, q_t, \Delta q_t, D_t] 
\end{align*}
\]
**Feature function:** \( \phi(a_{rank}, n_3) \)

**Query Features**
- Test if a search term \( w \in q_t \) and \( w \in q_{t-1} \)
- \# of times that a term \( w \) occurs in \( q_1, q_2, \ldots, q_t \)

**Query-Document Features**
- Test if a search term \( w \in +\Delta q_t \) and \( w \in D_{t-1} \)
- Test if a document \( d \) contains a term \( w \in -\Delta q_t \)
- \# of times \( tf-idf \) score of a document \( d \) to \( q_t \)

**Click Features**
- Test if there are SAT-Clicks in \( D_{t-1} \)
- \# of times a document being clicked in the current session
- \# of seconds a document being viewed and reviewed in the current session

**Query-Document-Click Features**
- Test if \( q_i \) leads to SAT-Clicks in \( D_i \), where \( i = 0 \ldots t-1 \)

**Session Features**
- position at the current session
## Experiment: Efficiency

<table>
<thead>
<tr>
<th>Approach</th>
<th>TREC 2012 Wall Clock</th>
<th>TREC 2012 CPU Cycle</th>
<th>Speed Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>winwin-long</td>
<td>$2.5 \times 10^4$ s</td>
<td>$6.8 \times 10^{13}$</td>
<td>1.00</td>
</tr>
<tr>
<td>winwin-short</td>
<td>$2.5 \times 10^4$ s</td>
<td>$6.8 \times 10^{13}$</td>
<td>1.00</td>
</tr>
<tr>
<td>qcm</td>
<td>$5.6 \times 10^3$ s</td>
<td>$1.5 \times 10^{13}$</td>
<td>4.46</td>
</tr>
<tr>
<td>dpl-upper</td>
<td>$3.2 \times 10^3$ s</td>
<td>$8.7 \times 10^{12}$</td>
<td>7.81</td>
</tr>
<tr>
<td>dpl</td>
<td>$3.1 \times 10^3$ s</td>
<td>$8.4 \times 10^{12}$</td>
<td>8.06</td>
</tr>
<tr>
<td>lemur</td>
<td>$2.6 \times 10^3$ s</td>
<td>$7.0 \times 10^{12}$</td>
<td>9.62</td>
</tr>
</tbody>
</table>

- *lemur > dpl > qcm > winwin* in Wall Clock and CPU Cycles
- We found similar conclusions on TREC 2013 and 2014 Session Track Experiments
**Experiment: Whole-Session Search Accuracy**

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>$\sigma^2$</th>
<th>MAP</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC median</td>
<td>0.2314</td>
<td>0.0575</td>
<td>0.1221</td>
<td>0.0152</td>
</tr>
<tr>
<td>TREC best (baseline)</td>
<td>0.3153</td>
<td>0.0844</td>
<td>0.1588</td>
<td>0.0191</td>
</tr>
<tr>
<td>lemur</td>
<td>0.2622</td>
<td>0.0800</td>
<td>0.1342</td>
<td>0.0160</td>
</tr>
<tr>
<td>winwin-short</td>
<td>0.2658</td>
<td>0.0785</td>
<td>0.1552</td>
<td>0.0195</td>
</tr>
<tr>
<td>qcm</td>
<td>0.3368$^\dagger$</td>
<td>0.0948</td>
<td>0.1536</td>
<td>0.0196</td>
</tr>
<tr>
<td>dpl</td>
<td>0.3382$^\dagger$</td>
<td>0.0932</td>
<td>0.1544</td>
<td>0.0193</td>
</tr>
<tr>
<td>winwin-long</td>
<td>0.3631$^\dagger$</td>
<td>0.0847</td>
<td>0.1651</td>
<td>0.0198</td>
</tr>
<tr>
<td>dpl-upper</td>
<td>0.3643$^\dagger$</td>
<td>0.0944</td>
<td>0.1609</td>
<td>0.0197</td>
</tr>
</tbody>
</table>

- **dpl** achieves a significant nDCG@10 improvement over the TREC best run
- **qcm < dpl < winwin-long < dpl-upper** in nDCG@10
- **dpl** achieves a good balance between accuracy and efficiency
OUTLINE

- A POMDP framework for modeling dynamic search
- Detecting user struggle state in search
- A direct policy learning framework
- Implementation of a dynamic search tool
IMPLEMENT A DYNAMIC SEARCH SYSTEM

○ TREC Dynamic Domain Track

○ Annotation Task:
  ● discover topics
  ● relevant passages for these topics

○ No pooling technology
  ● find a complete set of relevant passages and documents
<table>
<thead>
<tr>
<th>Searched for: availability of drugs</th>
<th>Lemur</th>
<th>soir</th>
<th>terrier</th>
<th>Results 1 - 50 of 600 matches; 1 tagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>ebola-a551658b8600b59193359e294800e475fdd765406f7ad09b205ef4615ded55</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make investigational Ebola drugs available, say ethicists</td>
<td>News</td>
<td>Pharmaceutical Journal</td>
<td>Home and News Analysis</td>
<td>Ebola drugs available</td>
</tr>
<tr>
<td>2. <strong>ebola-eff46b4e8546ce215416120326f5b6be04192d8d3207735169c08290974da656ba6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ebola.com &gt; drugs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.ebola.com">http://www.ebola.com</a> Ebola Information and Resources Fri, 09 Jan...drugs-treat-ebola/</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.ebola.com/?p=724">http://www.ebola.com/?p=724</a> (Time-Alexandra Sifferlin)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientists have identified 53 existing drugs that could be effective in fighting Ebola, according to newly published research. There is currently no vaccine or drug available to treat the disease, which is one of the primary...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. <strong>ebola-add38feb2bd9a815481b9a540d68870d7b4e06c24f7471d5852d7fabe38a7a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...Oncology Ophthalmic Disorders Oral Health Otic Disorders Pain Management Poisoning &amp; Drug Dependence Psychiatric Disorders Respiratory Disorders Urological Disorders Drug Interactions News Drug News Safety Alerts and Recalls New Products...Dependence Psychiatric Disorders Respiratory Disorders Urological Disorders Drug Interactions News Drug News Safety Alerts and Recalls New Products Drugs in the...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. <strong>ebola-da37cb6d6b4ad1b549ec622d09a6908765e4a9a47999a9c26b2cd1</strong> tagged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posted By AP On August 11, 2014 @ 8:57 am In Daily...issues, Spain announced that it had no vaccines or drugs available of the drug</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>published.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Health Ministry statement came less than a week after the U.S. Centers for Disease Control and Prevention said there were virtually no doses available of the drug that was used to treat two Americans with the disease. The drug...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. <strong>ebola-573ae332d70e89caafbb6fbf3c8e8b827a4a20df62d29a32ab5a34ab6d6fbf6f3b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make investigational Ebola drugs available, say ethicists</td>
<td>News</td>
<td>Pharmaceutical Journal</td>
<td>Home and News Analysis</td>
<td>Ebola drugs available</td>
</tr>
</tbody>
</table>
The ethical questions surrounding experimental Ebola drugs and vaccines were being debated Monday during a teleconference of medical ethicists and other experts organized by the U.N. health agency.

From doc: ebola-da37bed6b4ad1b54f9e622bd09ae6bcdb8...  
- marginally relevant - relevant - highly relevant - key result

The ethical questions surrounding experimental Ebola drugs and vaccines were being debated Monday during a teleconference of medical ethicists and other experts organized by the U.N. health agency.

From doc: ebola-da37bed6b4ad1b54f9e622bd09ae6bcdb8...  
- marginally relevant - relevant - highly relevant - key result
Nigerian officials say they had asked U.S. health authorities about getting the Ebola drug but were apparently not helped.

There is no known cure or licensed treatment for Ebola, which has killed over 960 people in the current outbreak in West Africa. The World Health Organization has called the Ebola outbreak — which emerged in Guinea in March and has since spread to Liberia, Sierra Leone and possibly Nigeria — an international health emergency and urged nations worldwide to battle the disease.

The ethical questions surrounding experimental Ebola drugs and vaccines were being debated Monday during a teleconference of medical ethicists and other experts organized by the U.N. health agency.

In a statement provided Monday, the Spanish Health Ministry said the ZMapp drug was obtained in Geneva this weekend with permission from the company and brought to Madrid to treat Miguel Pajares. The 75-year-old priest was evacuated from Liberia and placed in isolation Thursday at Madrid’s Carlos III Hospital.

Two Americans diagnosed with Ebola in Liberia and evacuated back to the United States have been treated with the drug. One of them, Dr. Kent Brantly, said last week that his condition was improving and the husband of the aid worker being
5 alternative search strategies

- three default search algorithms by lemur, solr and terrier
- one relevance feedback algorithm
- one dynamic search algorithm
Work to Complete

- Implement the dynamic search algorithm

Evaluation

- Check the mean and variance of number of relevant documents and passages found for the topics and subtopics.

- Check whether the ground truth can provide fair comparison to TREC Dynamic Domain Track submissions.
**Timeline**

<table>
<thead>
<tr>
<th>Deadline</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 18th, 2017</td>
<td>Dissertation proposal defense</td>
</tr>
<tr>
<td>End of Jul, 2017</td>
<td>Modify my dissertation and submit it to my committee members</td>
</tr>
<tr>
<td>End of Aug, 2017</td>
<td>Modify my dissertation based on committee feedback</td>
</tr>
<tr>
<td>End of Sep, 2017</td>
<td>Dissertation defense</td>
</tr>
<tr>
<td>Mid of Oct, 2017</td>
<td>Final modification and submit the dissertation to the graduate school</td>
</tr>
</tbody>
</table>
CONCLUSIONS

- Model dynamic search as a stochastic game by using a POMDP framework.

- View the search engine as an autonomous agent, that works together with user to collaborate on a shared task - fulfilling the information need.

- Provide a supervised learning model to infer searcher struggle state.

- Provide a direct policy learning model to improve efficiency.

- Hope our work calls for future adventures in the fields of POMDP in IR and game theory in IR.
**Future Work**

- Include other decision making states
  - user desire of revisiting webpages
  - personal interest and knowledge background
  - etc.

- Include more forms of user feedback to improve accuracy of inferring decision making states
  - implicit/explicit feedback
  - feedback from physical sensors

- Examine other utility functions besides clicks and nDCG as reward functions
  - optimize other aspects of search
    - e.g. maximize the match between retrieval content and user’s reading proficiency
    - minimize the browsing time, etc.
THANK YOU

Jiyun Luo

Department of Computer Science
Georgetown University
Washington, DC, USA
jl1749@georgetown.edu
The Win-win Search Framework

- A dual-agent stochastic game, \(<S, A_u, A_{se}, \Sigma_u, \Sigma_{se}, \Omega_u, \Omega_{se}, O, B, T, R>\).

<table>
<thead>
<tr>
<th>Name</th>
<th>symbol</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>(S)</td>
<td>the four hidden decision states</td>
</tr>
<tr>
<td>user action</td>
<td>(A_u)</td>
<td>add/remove/keep query terms</td>
</tr>
<tr>
<td>search engine action</td>
<td>(A_{se})</td>
<td>increase/decrease/keep term weights, adjust search techniques, etc.</td>
</tr>
<tr>
<td>message from user to search engine</td>
<td>(\Sigma_u)</td>
<td>clicked and SAT clicked documents</td>
</tr>
<tr>
<td>message from search engine to user</td>
<td>(\Sigma_{se})</td>
<td>top k returned documents</td>
</tr>
<tr>
<td>user's observation</td>
<td>(\Omega_u)</td>
<td>observations that the user makes from the world</td>
</tr>
<tr>
<td>search engine's observation</td>
<td>(\Omega_{se})</td>
<td>observations that the search engine makes from the world and from the user</td>
</tr>
<tr>
<td>user reward</td>
<td>(R_u)</td>
<td>relevant information the user gains from reading the documents</td>
</tr>
<tr>
<td>search engine reward</td>
<td>(R_{se})</td>
<td>nDCG that the search gains by returning documents</td>
</tr>
<tr>
<td>belief state</td>
<td>(B)</td>
<td>belief states generated from the belief updater and shared by both agents</td>
</tr>
</tbody>
</table>
DATA PREPARE: EXTRACT SEARCH TASKS

- extract cross-session search tasks from one-week-long Yahoo! search logs [Han et al. CIKM’16]

- sampling search tasks for annotation
DATA ANNOTATION

- Two assessors annotate each task independently
- A third assessor joins to solve disagreement
- Agree ratio: 67.4%
- Also check whether the search task contains multiple search topics
  - if yes, filter out

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>#Tasks</th>
<th>#Struggle</th>
<th>#Engaged</th>
<th>#Query/Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile</td>
<td>Mar. 9 ~ 15, 2015</td>
<td>1,226</td>
<td>483</td>
<td>743</td>
<td>5.58</td>
</tr>
<tr>
<td>Desktop</td>
<td></td>
<td>365</td>
<td>170</td>
<td>195</td>
<td>5.06</td>
</tr>
</tbody>
</table>
POMDP Components

- Observations

  the observation function \( O(s_j, a_t, \omega_t) \), defined as \( P(\omega_t | s_j, a_t) \), is the probability of observing \( \omega_t \in \Omega \) when agents take action \( a_t \) and land on state \( s_j \).

- Relevant or Non-relevant

  \( O(s_t = \text{Rel}, \Sigma_u, \omega_t = \text{Rel}) \propto P(s_t = \text{Rel} | \omega_t = \text{Rel})P(\omega_t = \text{Rel} | \Sigma_u) \)

  \( s_t \) is likely to be \( \begin{cases} \text{Relevant} & \text{if } \exists d \in D_{t-1} \text{ and } d \text{ is SATClicked} \\ \text{Non- Relevant} & \text{otherwise.} \end{cases} \)

  \( P(s_t = \text{Rel} | \omega_t = \text{Rel}) = \frac{\text{# of observed true relevant}}{\text{# of observed relevant}} \)

  “# of observed true relevant” is the number of times where the previously returned document set \( D_{t-1} \) contain at least one SAT clicks and those SAT clicked documents are indeed relevant documents in the ground truth.

  “# of observed relevant" is the number of times where \( D_{t-1} \) contains at least one SAT clicks.
POMDP Components

- Observations
  - Exploration or Exploitation

\[ O(s_t = \text{Exploration}, a_u = \Delta q_t, \Sigma_{se} = D_{t-1}, \omega_t = \text{Exploration} ) \]

\[ \propto P(s_t = \text{Exploration}|\omega_t = \text{Exploration}) \times P(\omega_t = \text{Exploration}|\Delta q_t, D_{t-1}) \]

- \( s_t \) is likely to be

\[ \begin{aligned}
\text{Exploration} & \quad \text{if } (+\Delta q_t \neq \emptyset \text{ and } +\Delta q_t \notin D_{t-1}) \\
& \quad \text{or } (+\Delta q_t = \emptyset \text{ and } -\Delta q_t \neq \emptyset) \\
\text{Exploitation} & \quad \text{if } (+\Delta q_t \neq \emptyset \text{ and } +\Delta q_t \in D_{t-1}) \\
& \quad \text{or } (+\Delta q_t = \emptyset \text{ and } -\Delta q_t = \emptyset) \\
\end{aligned} \]

\[ P(s_t = \text{Exploration}|\omega_t = \text{Exploration}) = \]

\[ \frac{\text{# of observed true explorations}}{\text{# of observed explorations}} \]

\[ P(\omega_t = \text{Exploration}|a = 'delete\_term', \Delta q_t, D_{t-1}) = \]

\[ \frac{\text{# of observed explorations due to deleting terms}}{\text{# of observations of deleting terms}} \]
BELIEF UPDATES

○ At every search iteration the belief state $b$ is updated when a new observation is obtained.

The belief update function is

$$b_{t+1}(s_j) = P(s_j | \omega_t, a_t, b_t)$$

$$= \frac{P(\omega_t | s_j, a_t, b_t) \sum_{s_i \in S} P(s_j | s_i, a_t, b_t) b_t(s_i)}{P(\omega_t | a_t, b_t)}$$

$$= O(s_j, a_t, \omega_t) \quad T(s_i, a_t, s_j)$$

$$T(s_i, a_t, s_j) = \frac{\#\text{Transition}(s_i, a_t, s_j)}{\#\text{Transition}(s_i, a_t, s^*)}$$

○ $\#\text{Transition}(s_i, a_t, s_j)$ is the number of all transitions that starts at state $s_i$, takes action $a_t$, and lands at state $s_j$.

○ $\#\text{Transition}(s_i, a_t, s^*)$ is the number of all transitions that starts at state $s_i$ and lands at any state by action $a_t$. 
JOINT OPTIMIZATION AND RETRIEVAL

○ The long term reward function for the search engine agent

\[ Q_{se}(b, a) = \sum_{s \in S} b(s) R(s, a) + \gamma \sum_{\omega \in \Omega} P(\omega | b, a_u, \Sigma_{se}) P(\omega | b, \Sigma_u) \max_a Q_{se}(b', a) \]

○ The long term reward function for the user agent

\[ Q_u(b, a_u) = R(s, a_u) + \gamma \sum a_u T(s_t | s_{t-1}, D_{t-1}) \max_{s_{t-1}} Q_u(s_{t-1}, a_u) \]

○ Joint optimization

\[ a_{se} = \arg\max_a (Q_{se}(b, a) + Q_u(b, a_u)) \]
**Search Accuracy**

- Search accuracy on TREC 2012 and 2013 Session

<table>
<thead>
<tr>
<th>Approach</th>
<th>nDCG@10</th>
<th>nDCG</th>
<th>MAP</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemur</td>
<td>0.2474</td>
<td>0.2627</td>
<td>0.1274</td>
<td>0.2857</td>
</tr>
<tr>
<td>TREC median</td>
<td>0.2608</td>
<td>0.2468</td>
<td>0.1440</td>
<td>0.2626</td>
</tr>
<tr>
<td>TREC best</td>
<td>0.3221</td>
<td>0.2865</td>
<td>0.1559</td>
<td>0.3595</td>
</tr>
<tr>
<td>PRF</td>
<td>0.2074</td>
<td>0.2335</td>
<td>0.1065</td>
<td>0.2415</td>
</tr>
<tr>
<td>Rocchio</td>
<td>0.2446</td>
<td>0.2714</td>
<td>0.1281</td>
<td>0.2950</td>
</tr>
<tr>
<td>Rocchio-CLK</td>
<td>0.2916†</td>
<td>0.2866</td>
<td>0.1449</td>
<td>0.3366</td>
</tr>
<tr>
<td>Rocchio-SAT</td>
<td>0.2889</td>
<td>0.2836</td>
<td>0.1467</td>
<td>0.3254</td>
</tr>
<tr>
<td>QCM+DUP</td>
<td>0.2742</td>
<td>0.2560</td>
<td>0.1537†</td>
<td>0.3221</td>
</tr>
<tr>
<td>QCM SAT</td>
<td>0.3350†</td>
<td>0.3054</td>
<td>0.1534†</td>
<td>0.1534</td>
</tr>
<tr>
<td>Win-Win</td>
<td>0.2941†</td>
<td>0.2691</td>
<td>0.1346</td>
<td>0.3403</td>
</tr>
</tbody>
</table>

- In TREC 2013, many relevant CatA documents are not included in the CatB collection, hence all systems' accuracy scores are lower than scores in TREC 2012.
Conclusions

- This paper presents a novel session search framework, win-win search, that uses a dual-agent stochastic game to model the interactions between user and search engine.

- This paper views the search engine as an autonomous agent, that works together with user to collaborate on a shared task - fulfilling the information needs.

- With a careful design of states, actions, and observations, the new framework is able to perform efficient optimization over a finite discrete set of options.

- We hope our work calls for future adventures in the fields of POMDP in IR and game theory in IR.

Acknowledgments: The author of this research is supported by SIGIR2014 student travel grant.
# Importance Features

<table>
<thead>
<tr>
<th>Mobile</th>
<th>Gain</th>
<th>PC</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvgQSatCTR</td>
<td>1.0000</td>
<td>AvgQClickFreq</td>
<td>1.0000</td>
</tr>
<tr>
<td>NumUniQuery</td>
<td>0.9420</td>
<td>AvgAddTerm</td>
<td>0.4772</td>
</tr>
<tr>
<td>AvgTerm</td>
<td>0.8649</td>
<td>RevisitClickRatio</td>
<td>0.4581</td>
</tr>
<tr>
<td>AvgQueryFreq</td>
<td>0.8445</td>
<td>AvgQSatCTR</td>
<td>0.4191</td>
</tr>
<tr>
<td>AvgQDwellTime</td>
<td>0.8118</td>
<td>AvgClickFreq</td>
<td>0.3684</td>
</tr>
<tr>
<td>AvgChar</td>
<td>0.7154</td>
<td>AvgChar</td>
<td>0.3376</td>
</tr>
<tr>
<td>AbandonRatio</td>
<td>0.6997</td>
<td>AvgQueryFreq</td>
<td>0.3352</td>
</tr>
<tr>
<td>AvgCDwellTime</td>
<td>0.5134</td>
<td>AvgQueryCTR</td>
<td>0.2923</td>
</tr>
<tr>
<td>AvgQueryCTR</td>
<td>0.5081</td>
<td>DurationSAT_1</td>
<td>0.2220</td>
</tr>
<tr>
<td>TotalClickDwell</td>
<td>0.4735</td>
<td>AvgQueryEdit</td>
<td>0.1942</td>
</tr>
</tbody>
</table>
### Feature Importance Summary

<table>
<thead>
<tr>
<th>Mobile only</th>
<th>PC only</th>
<th>Both platforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>NumUniQuery +</td>
<td>AvgAddTerm +</td>
<td>AvgQSatCTR −</td>
</tr>
<tr>
<td>AvgQDwellTime −</td>
<td>RevisitClickRatio −</td>
<td>AvgQClickFreq −</td>
</tr>
<tr>
<td>AbandonRatio +</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgTerm +</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvgQueryFreq −</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Typing many unique queries is a good indicator of struggles in mobile search only.
- Task complexity and material inconsistency may cause user struggles more easily in search using PCs.
- Writing/formulating long queries, using lower frequency queries and clicking less popular URLs are good indicators of struggles on both platforms.

* +/- indicates the signs of the features’ coefficients (the correlation with the struggling label) in the LM models.
**Data Prepare: Extract Search Tasks**

- extract cross-session search tasks from one-week-long Yahoo! search logs
  - for all query pairs \((q_i, q_j)\) generated by the same user
    - a logistic regression model to predict whether they are topic-coherent [Han at al. CIKM’16]
  - search task clustering
    - given \((q_i, q_j)\) are topic-coherent, if \(q_i/q_j\) belongs to a task cluster, then merge them into the same task cluster
- sampling search tasks for annotation
Q1: quit smoking

D1: Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help

…

Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction

…

Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...
Example: TREC 2014 Session 1011
“quit smoking”

Q1  quit smoking
D1  Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
    Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ...
    Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1  [clicked:none, q1, +Δq:quit smoking, −Δq:none, D1]
Example: TREC 2014 Session 1011
“quit smoking”

q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help …
  Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction …
  Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs…

h1 [clicked: none, q1, +Δq:quit smoking, −Δq:none, D1]
C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help …
  Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction …
  SAT-Clicked. Dwell time: 24 seconds
  Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs… Dwell time: 40 seconds
Q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
... Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ...
... Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1 [clicked:none, q1, +Δq:quit smoking, −Δq:none, D1]
C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
... Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction Dwell time: 24 seconds
... Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs... SAT-Clicked.

Q2 smoking quitting hypnosis
D2 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
... Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...
Example: TREC 2014 Session 1011
"quit smoking"

q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
   Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ...
   Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

[clicked: none, q1, +Δq: quit smoking, -Δq: none, D1]

C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
   Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ... SAT-Clicked.
   Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...  Dwell time: 40 seconds

Q2 smoking quitting hypnosis
D2 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
   ... Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...

Query reformulation using words in previous search results
Example: TREC 2014 Session 1011
“quit smoking”

q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help...
... Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction...
Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1 [clicked:none, q1, +Δq:quit smoking, −Δq:none, D1]
C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help...
... Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction... Dwell time: 24 seconds
SAT-Clicked.
Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs... Dwell time: 40 seconds

q2 smoking quitting hypnosis
D2 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
... Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...

h2 [h1, clicked: [[3, 24, SAT-Clicked=F], [6, 40, SAT-Clicked=T]], q2, +Δq:hypnosis, −Δq:none, D2]
Example: TREC 2014 Session 1011
“quit smoking”

q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
  Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction ...
  Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1 [clicked: none, q1, +Δq: quit smoking, −Δq: none, D1]
C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
  Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction ...
  Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

q2 smoking quitting hypnosis
D2 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
  Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...

h2 [h1, clicked: [[3, 24, SAT-Clicked=F], [6, 40, SAT-Clicked=T]], q2, +Δq: hypnosis, −Δq: none, D2]
C3 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
  Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart... SAT-Clicked. Dwell time: 24 seconds

q3 side effects quit smoking hypnosis
D3 Rank 1: Side Effects Of Quitting Smoking | Self Hypnosis To Quit Smoking ...
Example: TREC 2014 Session 1011
“quit smoking”

```
q1  quit smoking
D1  Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
    Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ...
    Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1  [clicked:none, q1, +Δq:quit smoking, -Δq:none, D1]
C2  Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
    ... SAT-Clicked.
    Rank 3: Quit Smoking Toolbox - Quit Smoking - Nicotine Addiction ... Dwell time:24 seconds
    ... Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs... SAT-Clicked.
q2  smoking quitting hypnosis
D2  Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking
    Hypnosis Neuro...
    ... Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...

h2  [h1, clicked:[[3,24,SAT-Clicked=F],[6,40,SAT-Clicked=T]],q2,+Δq:hypnosis, -Δq:none, D2 ]
C3  Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking
    Hypnosis Neuro...
    ... Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...
        SAT-Clicked.
        Dwell time: 40 seconds
q3  side effects quit smoking hypnosis
D3  Rank 1: Side Effects Of Quitting Smoking | Self Hypnosis To Quit Smoking ...
Example: TREC 2014 Session 1011
“quit smoking”

q1 quit smoking
D1 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
…
Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction ...
…
Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs...

h1 [clicked: none, q1, +Δq: quit smoking, −Δq: none, D1]
C2 Rank 1: Easy Ways to Quit Smoking | Quit Smoking Help ...
…
Rank 3: Quit Smoking Toolbox – Quit Smoking – Nicotine Addiction … Dwell time: 24 seconds
…
Rank 6: Quit Smoking Hypnosis, Stop Smoking Hypnosis CDs… SAT-Clicked.
Dwell time: 40 seconds

q2 smoking quitting hypnosis
D2 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
…
Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart...

h2 [h1, clicked:[[3,24,SAT-Clicked=F],[6,40,SAT-Clicked=T]],q2,+Δq:hypnosis, −Δq: none, D2 ]
C3 Rank 1: Quit Smoking Hypnosis | Stop Smoking Hypnosis CDs Quit Smoking Hypnosis Neuro...
…
Rank 4: Quit Smoking with Video Hypnosis Home Shopping Cart… SAT-Clicked.
Dwell time: 31 seconds

q3 side effects quit smoking hypnosis
D3 Rank 1: Side Effects Of Quitting Smoking | Self Hypnosis To Quit Smoking …
…

h3 [h2, clicked:[[4,31,SAT-Clicked=T]],q3,+Δq: side effects, −Δq: hypnosis, D3 ]
Decompose a history

• First level: iteration by iteration

• Second level: break down an iteration into
  • browse phase
  • query phase
  • retrieval phase
Browse Phase

• Actor: the user

• It happens
  • after the search results are shown to the user
  • before the user starts to write the next query

• Records how the user perceives and examines the (previously retrieved) search results
Query Phase

• Actor: the user

• It happens
  • when the user writes a query

• Assuming the query is created based on
  • what has been seen in the browse phase

• the information need
Rank Phase

- Actor: the search engine
- It happens
  - after the query is entered
  - before the search results are returned
- It is where the search algorithm takes place

\[ s(t) \xrightarrow{\text{browse}} a_{\text{browse}}(t) \xrightarrow{\text{query}} a_{\text{query}}(t) \xrightarrow{\text{rank}} s(t+1) \]

\[ o_{\text{rank}}(t) \xrightarrow{\text{browse}} n_1(t) \xrightarrow{\text{query}} n_2(t) \xrightarrow{\text{rank}} n_3(t) \]
Our objective function:

\[
V_\theta(s_0) = \sum_{t=0}^{\infty} \sum_{h \in H} \gamma^t P(h|\theta) r(t, h)
\]

where

\[
P(h|\theta) = \prod_{t=1}^{\text{len}(h)} P(o_{\text{rank}}(t), a_{\text{browse}}(t),
\]

\[o_{\text{browse}}(t), a_{\text{query}}(t), o_{\text{query}}(t), a_{\text{rank}}(t)|h_{t-1}, \theta)
\]

\[
\propto \prod_{t=1}^{\text{len}(h)} P(a_{\text{browse}}(t)|o_{\text{rank}}(t), \theta_1)
\]

\[\times P(a_{\text{query}}(t)|o_{\text{browse}}(t), \theta_2)
\]

\[\times P(a_{\text{rank}}(t)|o_{\text{browse}}(t), o_{\text{query}}(t), o_{\text{rank}}(t), \theta_3)
\]

\[
\propto \prod_{t=1}^{\text{len}(h)} \prod_{i \in \{1,2,3\}} P(a^i(t)|n_i(t), \theta_i)
\]
Action Selection Distribution

Softmax Function \( P(a^i|n_i, \theta_i) = \frac{e^{\theta_i \cdot \phi(a^i, n_i)}}{\sum_{a'^i} e^{\theta_i \cdot \phi(a'^i, n_i)}} \)

Gradient

\[
\frac{\partial V_{\theta}(s_0)}{\partial \theta_k} = \sum_{t=1}^{\infty} \gamma^t \sum_{h \in H} r(t, h) \frac{\partial P(h|\theta)}{\partial \theta_k} \\
= \sum_{t=1}^{\infty} \gamma^t \sum_{h \in H} r(t, h) P(h|\theta) \\
\times \sum_{i=0}^{t} \frac{\partial \ln[P(a_{\text{browse}}|n_1, \theta_1)P(a_{\text{query}}|n_2, \theta_2)P(a_{\text{rank}}|n_3, \theta_3)]}{\partial \theta_k}
\]
Ranking Function

\[ P(a_{rank}|n_3, \theta_3) = \frac{e^{\theta_3 \cdot \phi(a_{rank}, n_3)}}{\sum_{a'_{rank}} e^{\theta_3 \cdot \phi(a'_{rank}, n_3)}} \]

• It originally presents the probability of selecting a (ranking) action

• In our context, the probability of selecting d to be put at the top of a ranked list under \( n_3 \) and \( \theta_3 \) at the \( t^{\text{th}} \) iteration

• Then we sort the documents by it to generate the document list
**Algorithm 1** Direct Policy Learning for Session Search.

```
procedure \( \text{DPL}(H, \epsilon) \)
\( \triangleright \) \( H \) is the training history set. \( \epsilon \) is a threshold.
\( \theta_3 \leftarrow \text{random}(0, 1) \)

repeat
  Sample history \( h \) from \( H \):
  \( q_0, D_0, C_0, T_0 \leftarrow \emptyset, r(1, h) \leftarrow 0, \Delta \theta_3 \leftarrow 0 \)

  for \( t = 1 \) to \( \text{len}(h) \) do
    \( o_{\text{rank}} \leftarrow D_{t-1}, n_1 \leftarrow o_{\text{rank}} \)
    \( \triangleright a_{\text{browse}} \) is performed by the user
    \( o_{\text{browse}} \leftarrow (C_t, T_t), n_2 \leftarrow o_{\text{browse}} \)
    \( r(t, h) \leftarrow \text{CalculateReward}(D'_t, o_{\text{browse}}, h) \)

    \( \triangleright a_{\text{query}} \) is performed by the user
    \( o_{\text{query}} \leftarrow \text{GetQueryChange}(q_t, q_{t-1}) \)
    \( n_3 \leftarrow (o_{\text{rank}}, o_{\text{browse}}, o_{\text{query}}) \)

    Sample a search engine action \( a_{\text{rank}} \)
    \( \sim P(a_{\text{rank}}|n_3, \theta_3) \)
    \( D'_t \leftarrow \text{DocRanking}(a_{\text{rank}}) \)

    \( \Delta \theta_3 \leftarrow \text{UpdateGradient}(r(t, h), D'_t, n_3, \theta_3) \)
    \( \theta_3 \leftarrow \theta_3 + \Delta \theta_3 \)

  end for

until \( \Delta \theta_3 < \epsilon \) or \( H \) is running out
```

**Updates:**
\[ \Delta \theta_3 = \sum_{h \in H} \sum_{t=1}^{\text{len}(h)} \gamma^t r(t, h) \times \sum_{i=1}^{t} [\phi(a_{\text{rank}}, n_3) - \sum_{a'_{\text{rank}}} \phi(a'_{\text{rank}}, n_3) P(a'_{\text{rank}}|n_3, \theta_3)] \]

**Feature function:** \( \phi(a_{\text{rank}}, n_3) \)

- **Query Features**
  - Test if a search term \( w \in q_t \) and \( w \in q_{t-1} \)
  - \# of times that a term \( w \) occurs in \( q_1, q_2, ..., q_t \)

- **Query-Document Features**
  - Test if a search term \( w \in +\Delta q_t \) and \( w \in D_{t-1} \)
  - Test if a document \( d \) contains a term \( w \in -\Delta q_t \)

- **Click Features**
  - Test if there are SAT-Clicks in \( D_{t-1} \)
  - \# of times a document being clicked in the current session
  - \# of seconds a document being viewed and reviewed in the current session

- **Query-Document-Click Features**
  - Test if \( q_i \) leads to SAT-Clicks in \( D_i \), where \( i = 0...t-1 \)

**Session Features**
- position at the current session