An Optimization Framework for Query Recommendation

Aris Anagnostopoulos, Luca Becchetti,
Carlos Castillo, Aristides Gionis
As you set out for Ithaca, hope your road is a long one, full of adventure, full of discovery.

K. Kavafis
Problem

Given a set of possible user histories
Problem

Given a value for different states
Problem

Given a value for different states
Problem

“Nudge” the users in a certain direction
“Nudge” the users in a certain direction
Problem definition

- Given a set of possible user sessions
- Given a certain value for different states
- “Nudge” the users in a certain direction

Objectives

- 1. collect a large reward along the way - or -
- 2. end the session at a rewarding action
Constraints

- We are not almighty

Source: not-of-this-earth.com
Constraints

• We are not almighty
  – We can only suggest, not order
Constraints

• We are not almighty
  – We can only suggest, not order

• We are not all-knowing

Source: Wikimedia commons.
Constraints

• We are not almighty
  – We can only suggest, not order

• We are not all-knowing
  – We do not know how the users will react
Framework
Setting

• Paper: general framework
  – e.g. for optimizing links on web sites

• Talk: query recommendation
Also try: central park zoo, central park new york, More...

Central Park Conservancy
Official site with history, things to do, virtual tour, and how to get involved.
www.centralparknyc.org - Cached

Central Park Conservancy: Central Park Website: Virtual Park
Would you like your own copy of the Central Park map, complete with details? ... To print a comprehensive map from a pdf file, click on Central Park map. ...
www.centralparknyc.org/site/PageNavigator/virtualpark_main - Cached

CentralPark.com | Your Complete Guide to New York's Central Park
Your first stop for information on Central Park New York City and The Central Park Zoo! Events, park attractions, activities, weddings, maps, tours, history, tourist guide
www.centralpark.com - Cached
Also try: central park zoo, central park new york, More...

Central Park Conservancy
Official site with history, things to do, virtual tour, and how to get involved.
www.centralparknyc.org - Cached

Central Park Conservancy: Central Park Website: Virtual Park
Would you like your own copy of the Central Park map, complete with details? ... To print a comprehensive map from a pdf file, click on Central Park map. ...
www.centralparknyc.org/site/PageNavigator/virtualpark_main - Cached

CentralPark.com | Your Complete Guide to New York's Central Park
Your first stop for information on Central Park New York City and The Central Park Zoo! Events, park attractions, activities, weddings, maps, tours, history, tourist guide
www.centralpark.com - Cached
Query recommendations

- Reformulation probabilities
  - $P(q,q')$ original

![Diagram showing reformulation probabilities for different queries related to Central Park in New York City.](image)
Query recommendations

- Reformulation probabilities
  - $P(q,q')$ original
  - $P'(q,q')$ perturbed = $P(q,q') + \rho(Q,q,q')$
Query recommendations

- Reformulation probabilities
  - $P(q,q')$ original
  - $P'(q,q')$ perturbed = $P(q,q') + \rho(Q,q,q')$

```
central park
\begin{itemize}
  \item 0.2 \hspace{1cm} central park map
  \item 0.1 \hspace{1cm} central park hotel
  \item 0.3 \hspace{1cm} central park new york
  \item 0.4 \hspace{1cm} central park zoo
\end{itemize}
```
Example values \( w(\cdot) \)

- Search engine results page
  - Quality of search results
- General page
  - Dwell time
  - User ratings
Objective functions $U(\cdot)$

**Kavafian**

"a road full of adventure"

$$U(q_1, q_2, \ldots, q_t) = \sum w(q_i)$$

**Machiavellian**

"ends justify means"

$$U(q_1, q_2, \ldots, q_t) = w(q_t)$$
The *Kavafian* objective

Useful when users want to explore, or be entertained
The *Kavafian* objective

Useful when users want to explore, or be entertained

"funny video"
Optimization problem

• Given:
  – Original transition probabilities $P$
  – Starting node $q$
  – Node values $w$
  – Perturbation function $\rho$

• Add up to $k$ links (per node), maximize expected utility of paths starting at $q$
Single-step recommendation

- Recommend now, leave user alone later
Single-step recommendation

• Recommend now, leave user alone later
Multi-step recommendation

- Recommend at each step in the future
Multi-step recommendation

- Recommend at each step in the future
Multi-step case

- **Multi- and Single-step** recommendation are NP-complete
  - Reduction from MAXIMUM-COVER

- Heuristic for multi-step problem:
  - at each node, assume rest of the graph is unperturbed when computing utility of adding an edge
Single-step case

• Greedy heuristic for “Machiavellian” objective: find \((q_i, q_j)\) maximizing

\[
p_{ij}(E[U(path(q_j))] - w_i)
\]

• Repeat k times
Observation

- Greedy heuristic achieves utility at least \((1-x)\) of the optimum, with \(x \ll 1\) in cases of practical interest
  - It is possible to construct pathological instances s.t. that greedy performs poorly
  - \(x\) depends on termination probability at \(q_i\) and probabilities of following recommendations
Application
Large-scale experiment

We observe perturbed probabilities $P'(q,q')$, unless we disable search assist to see $P(q,q')$.
Empirical observations

• Notation:
  – $\tau_q, \tau'_q$ session-end probabilities

• Recommendations decrease termination probability:
  · Average $\tau_q \approx 0.90$
  · Average $\tau'_q \approx 0.84$
Termination probability

\[ \tau_q \quad \text{and} \quad \tau'_q \]
Empirical observations

- Recommendations decrease termination probability, $\tau_q \approx 0.90$ $\tau'_q \approx 0.84$

- Decrease is almost entirely due to more clicks on recommendations
$r = 0.95$

The plot shows a scatter plot with a linear relationship between two variables. The correlation coefficient $r$ is 0.95, indicating a strong positive correlation. The equation $1.0\tau_q - 0.8\sum \rho_{q,q'}$ is plotted along the x-axis.
Empirical observations

• Recommendations decrease termination probability from $\approx 0.90$ to $\approx 0.84$

• Decrease is almost entirely due to more clicks on recommendations

• $\rho$ is difficult to estimate
$P(q, q')$: without recommendations

$P'(q, q')$: with recommendations
There is an increase in $P(q,q')$ with recommendations compared to $P(q,q')$: without recommendations.
Many un-seen suggestions are clicked when shown
The rest are in general difficult to predict.
So what do we do?

- We approximate ρ by a linear function on
  - P(q,q')
  - Textual similarity of q and q'
  - Terminal probability of q

- We have a low accuracy on this prediction
  - r ≈ 0.5

- We use as weights the CTR on results
Evaluation
Evaluation results

• Baselines:
  – Prefer queries with large $w$
  – Prefer queries with large $\rho$
  – Prefer queries with large $\rho w$

• Greedy heuristic performs better for both utility functions
Expected utility

“Kavafian” objective

“Machiavellian” objective
Evaluation results

- Greedy heuristic performs well
- What about relevance?
  - 420 queries assessed by 3 judges
  - There were no significant changes in relevance between systems
Conclusions

• General framework for "nudging" users in a certain direction

• Open algorithmic and practical questions

• In the paper: related work
Q&A

chato@yahoo-inc.com