User Browsing Models: Relevance versus Examination

Ramakrishnan Srikant
Sugato Basu
Ni Wang
Daryl Pregibon

Background

Estimating Relevance of Search Engine Results

- Use CTR (click-through rate) data.
- Pr(click) = Pr(examination) x Pr(click | examination)

Relevance

- Need user browsing models to estimate Pr(examination)

Notation

- Φ(i) : result at position i
- Examination event: $E_i = \begin{cases} 1, \text{if the user examined } \phi(\text{i}) \\ 0, \text{otherwise} \end{cases}$
- Click event: $C_i = \begin{cases} 1, \text{if the user clicked on } \phi(\text{i}) \\ 0, \text{otherwise} \end{cases}$

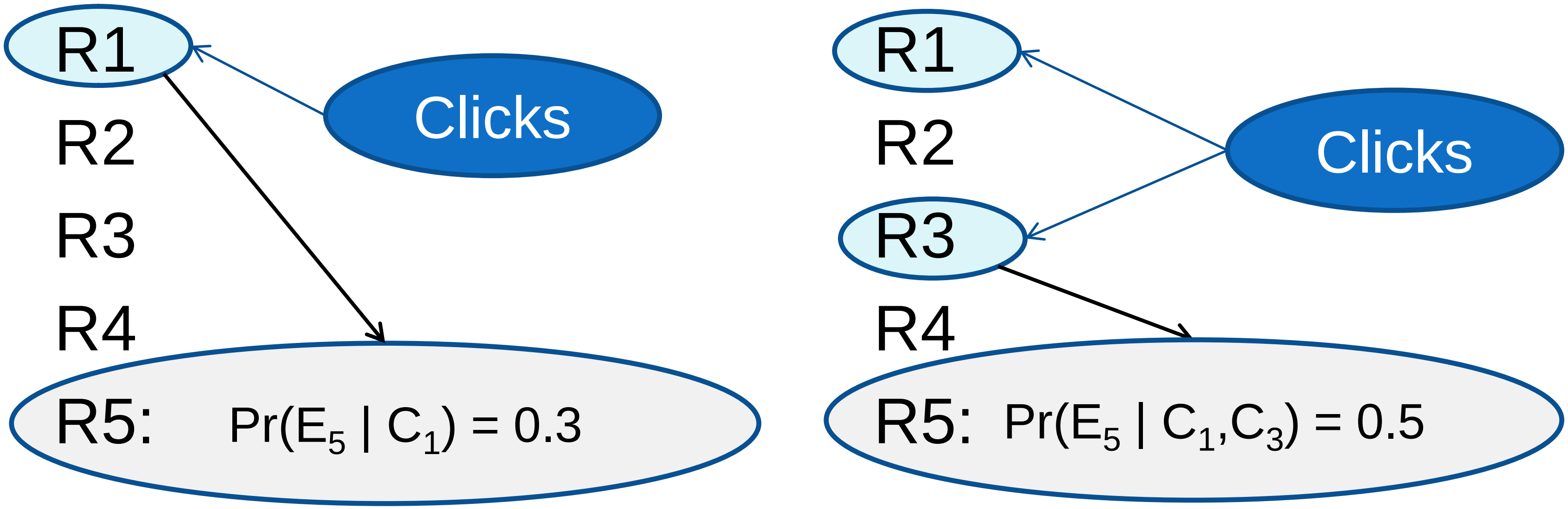
Examination Hypothesis

Richardson et al, WWW 2007:

$$\Pr(C_i = 1) = \Pr(E_i = 1) \; \Pr(C_i = 1 \mid E_i = 1)$$

- α_i : position bias
  - Depends solely on position.
  - Can be estimated by looking at CTR of the same result in different positions.

Using Prior Clicks



Examination depends on prior clicks

- Cascade model
- Dependent click model (DCM)
- User browsing model (UBM) [Dupret & Piwowarski, SIGIR 2008]
  - More general and more accurate than Cascade, DCM.
  - Conditions Pr(examination) on closest prior click.
- Bayesian browsing model (BBM) [Liu et al, KDD 2009]
  - Same user behavior model as UBM.
  - Uses Bayesian paradigm for relevance.

User browsing model (UBM)

- Use position of closest prior click to predict Pr(examination).

$$\Pr(E_i = 1 \mid C_{1:i-1}) = \alpha_i \, \beta_{i,p(i)}$$

position bias

p(i) = position of closest prior click

$$\Pr(C_i = 1 \mid C_{1:i-1}) = \Pr(E_i = 1 \mid C_{1:i-1}) \Pr(C_i = 1 \mid E_i = 1)$$

Prior clicks don't affect relevance.

Other Related Work

- Examination depends on prior clicks and prior relevance
  - Click chain model (CCM)
  - General click model (GCM)
- Post-click models
  - Dynamic Bayesian model
  - Session utility model

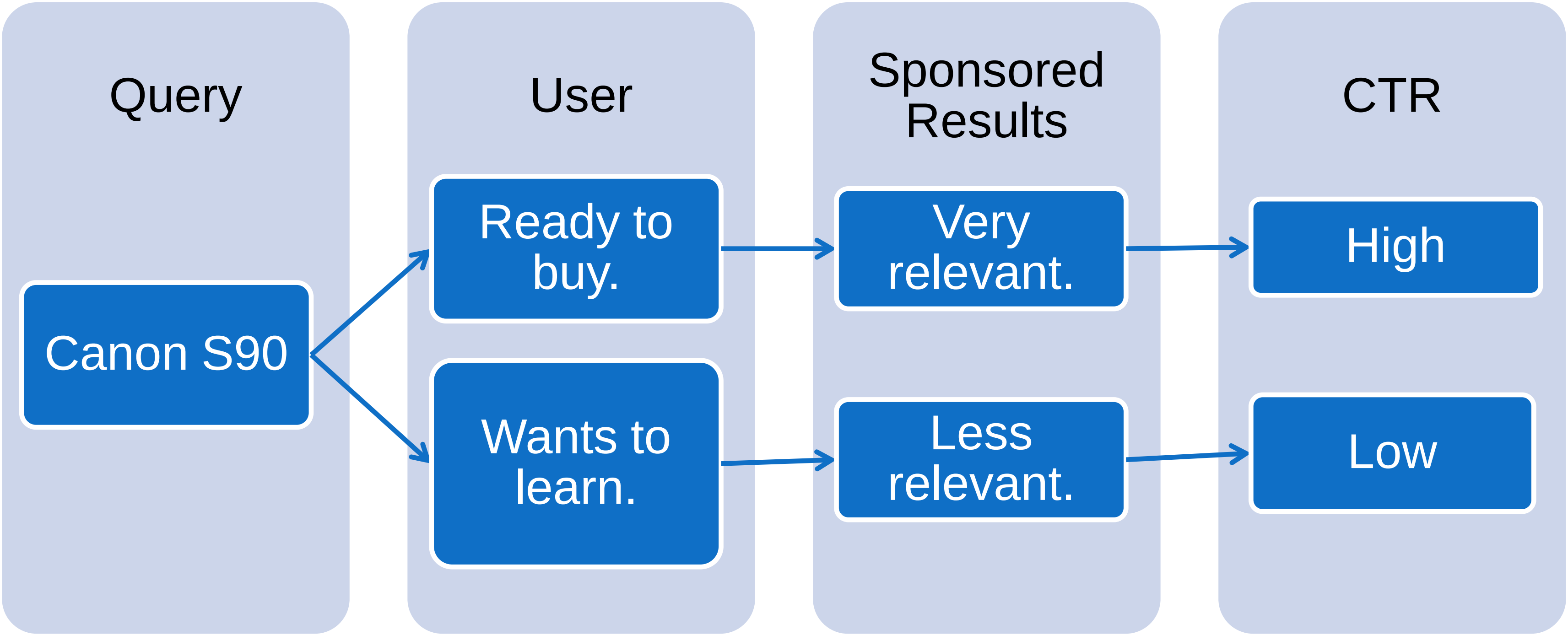
Constant Relevance Assumption

Constant Relevance Assumption

- Cascade model, DCM, UBM, BBM, CCM, GCM all implicitly assume:
  - Relevance is independent of prior clicks.
  - Relevance is constant across query instances.
    - Query = “Canon S90”
      - Aggregate relevance: Relevance to a query.
    - Query instance = “Canon S90” for a specific user at a specific point in time.
      - Instance relevance: Relevance to a query instance.

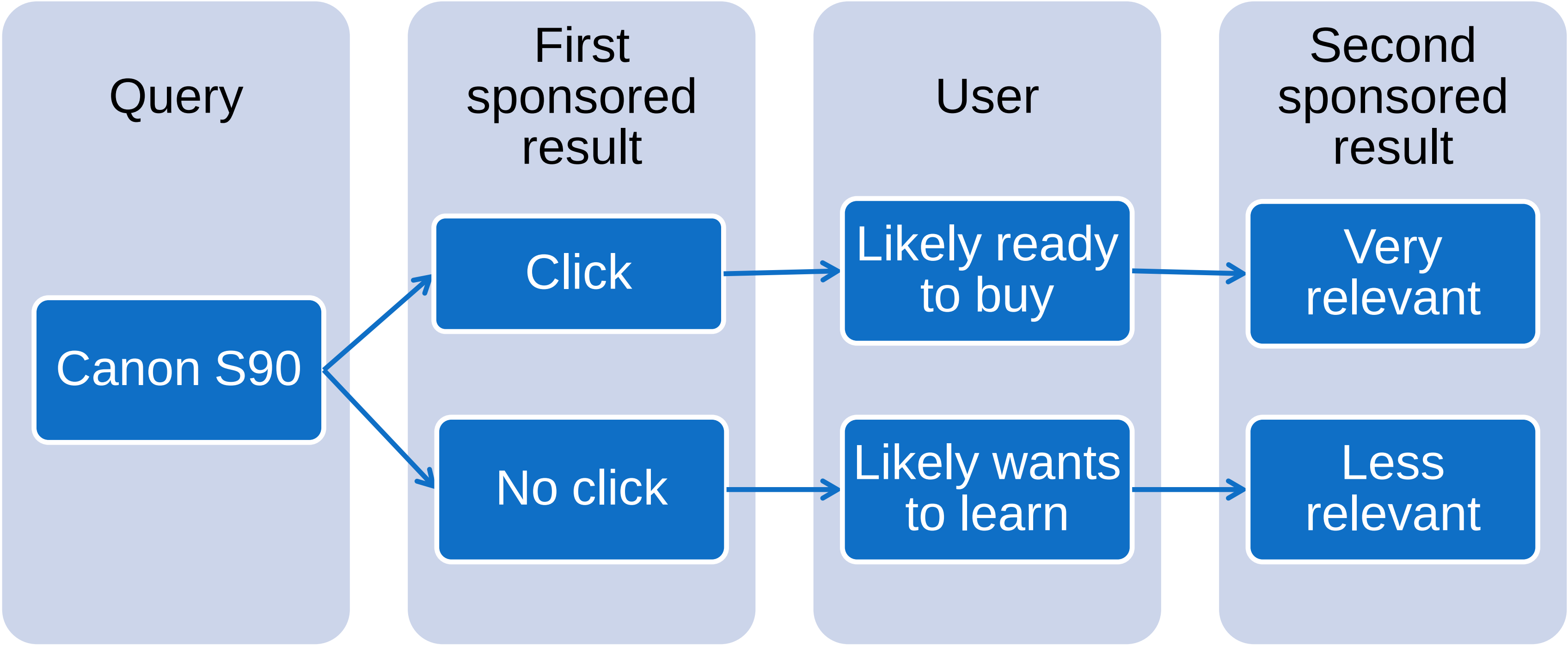
User intent

- Query string does not fully capture user intent.



Prior clicks signal relevance.

... not just Pr(examination).

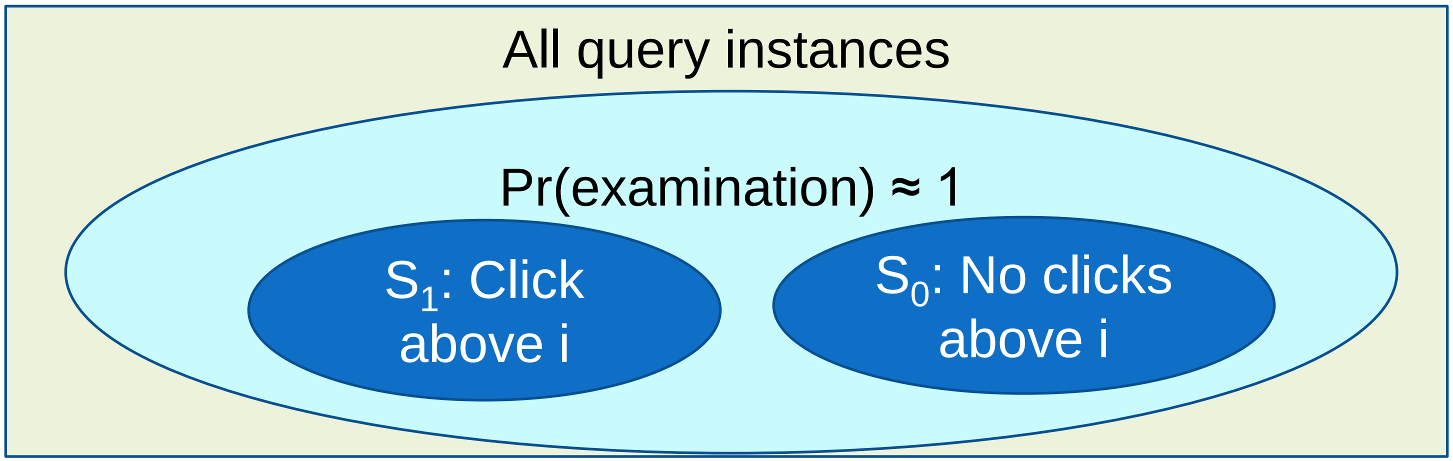


Testing Constant Relevance

- If we know that Pr(examination) ≈ 1:
  - Relevance ≈ CTR
  - Test whether relevance is independent of prior clicks.
- When is Pr(examination) ≈ 1?
  - Users scan from top to bottom.
  - If there is a click *below* position i, then $Pr(E_i) \approx 1$.

Statistical Power of the Test

- For a specific position i:



- If relevance is independent of prior clicks, we expect

$$\frac{\text{Clicks}(S_1)}{\text{Predicted clicks}(S_1)} \approx \frac{\text{Clicks}(S_0)}{\text{Predicted clicks}(S_0)}$$

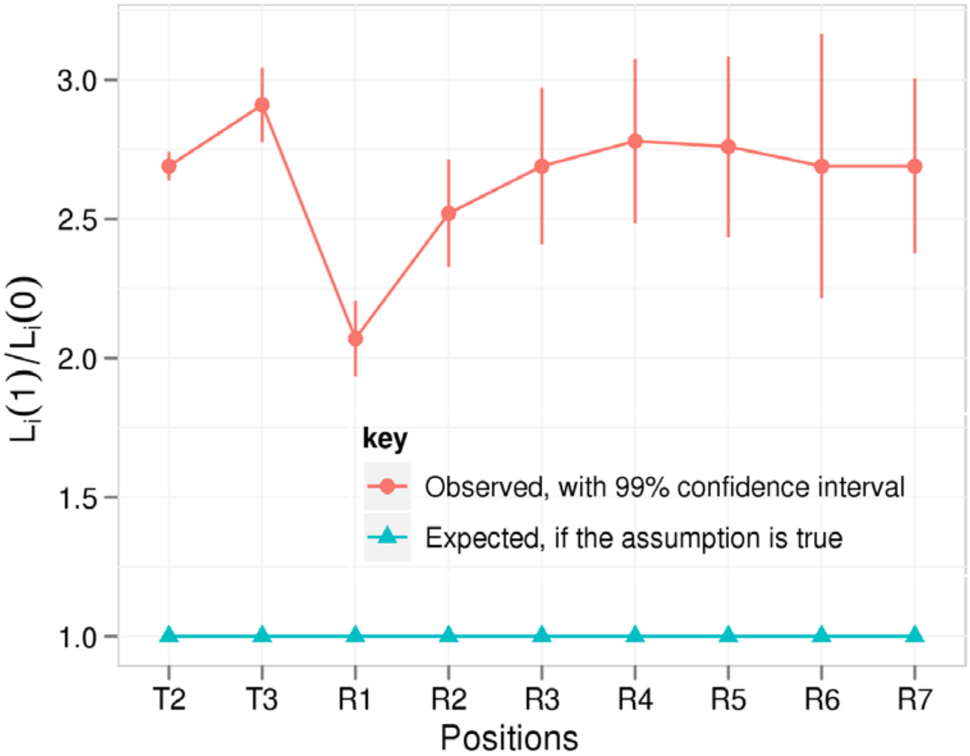
Lift = LHS / RHS ≈ 1

The data speaks…

- Over all configs:

  Lift = 2.69 +/- 0.05 (99% conf. Interval)

- Graph shows config with 3 top ads, 8 rhs ads.
- T2 = 2nd top ad, R3 = 3rd rhs ad, etc.



New User Browsing Models

Pure Relevance

Max Examination

Joint Relevance Examination (JRE)

Pure Relevance

- Any change in $Pr(C_i = 1)$ when conditioned on other clicks is solely due to change in instance relevance.
- Number of clicks on other results used as signal of instance relevance.
  - Does not use position of other clicks, only the count.
- Yields *identical* aggregate relevance estimates as the baseline model (which does not use co-click information).

$$Pr(C_i = 1 \mid C_{\neq i}, E_i = 1) = r_{\Phi(i)} \delta_{n(i)}$$

n(i) = number of click in other positions

Max Examination

- Like UBM/BBM, but also use information about clicks below position i.
  - Pr(examination) ≈ 1 if there is a click below i
- UBM/BBM: $\Pr(E_i = 1 \mid C_{1:i-1}) = \alpha_i \beta_{i,p(i)}$
- Max-examination: $\Pr(E_i = 1 \mid C_{\neq i}) = \alpha_i \beta_{i,e(i)}$

$$e(i) = \begin{cases} p(i), & \text{if no click below position i} \\ i+1, & \text{if there is a click below i} \end{cases}$$

p(i) = position of closest prior click

Joint Relevance Examination (JRE)

- Combines the features of the pure relevance and max-examination models.
- Allows CTR changes to be caused by both changes in examination and changes in instance relevance.

$$\Pr(E_i = 1 \mid C_{\neq i}) = \alpha_i \, \beta_{i,e(i)}$$

$$\Pr(C_i = 1 \mid C_{\neq i}, E_i = 1) = r_{\Phi(i)} \, \delta_{n(i)}$$

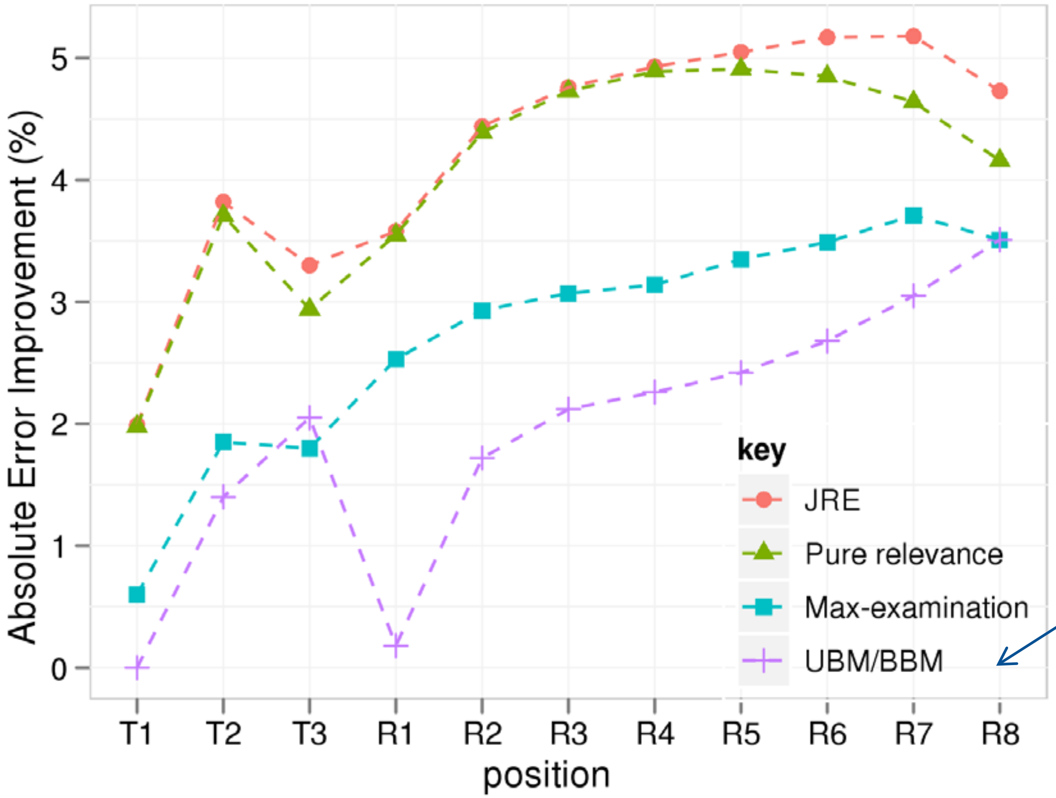
$$\Pr(C_i = 1 \mid C_{\neq i}) = \Pr(E_i = 1 \mid C_{\neq i}) \Pr(C_i = 1 \mid E_i = 1, C_{\neq i})$$

Predicting CTR

Predicting CTR

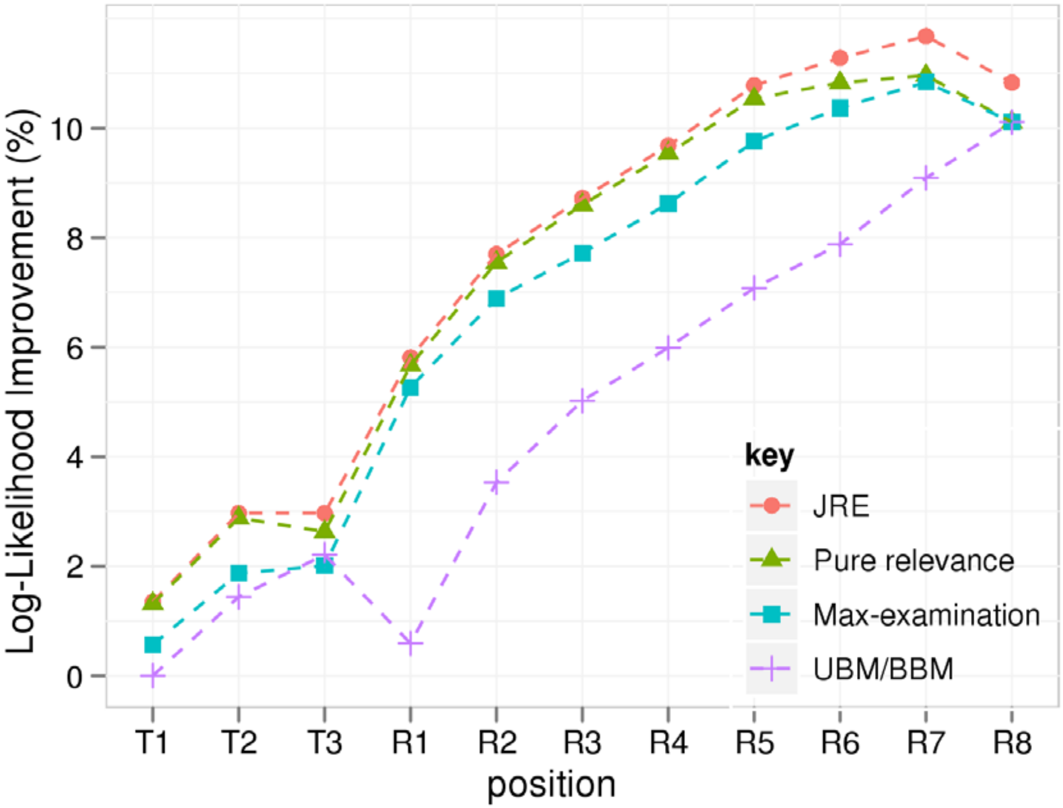
- Models:
  - Baseline: Google's production system for predicting relevance of sponsored results.
    - Does not use co-click information.
  - Compare to UBM/BBM, max examination, pure relevance, and JRE.
- Data:
  - 10% sample of a week of data.
  - 50-50 split between training and testing.

Absolute Error

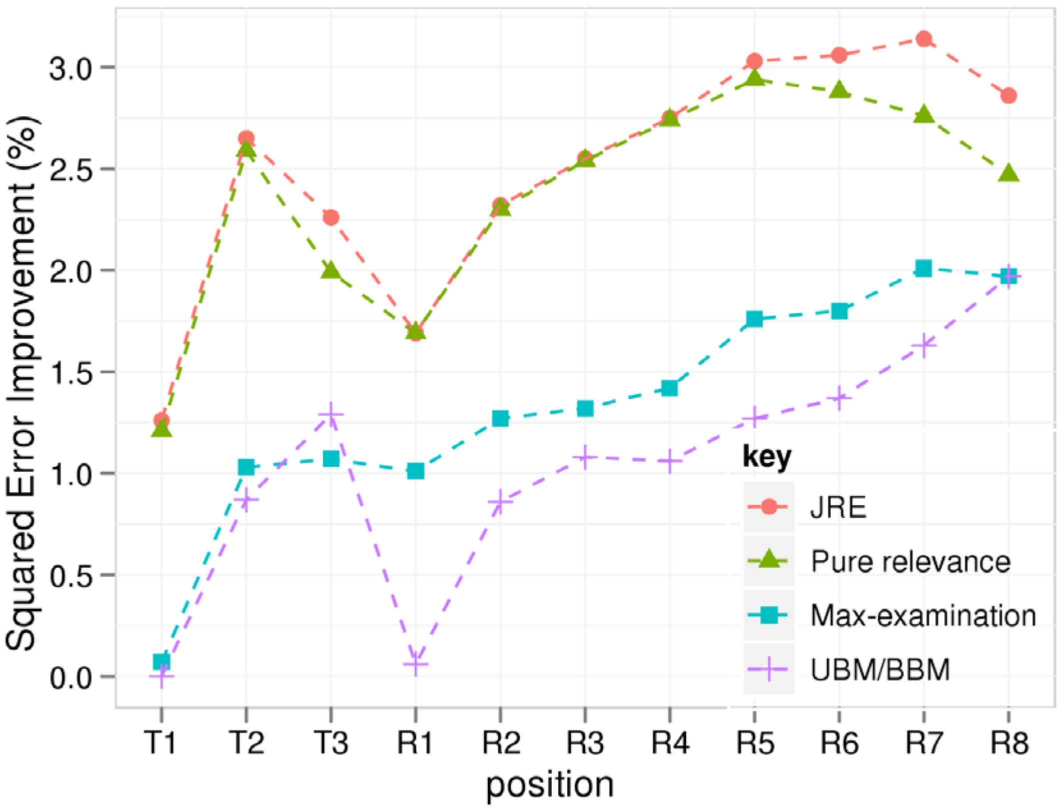


Baseline: Google's production system.

Log likelihood



Squared Error



Predicting Relevance

Predicting Relevance vs. Predicting CTR

- If model A is more accurate than model B at predicting CTR, wouldn't A also be better at predicting (aggregate) relevance?

Counter-Example

- CTR:

  pure-relevance
  >>
  max-examination
  >>
  baseline

- Relevance: Either
  - pure-relevance == baseline
    >>
    max-examination
    OR
  - max-examination
    >>
    pure-relevance == baseline

Intuition

- Predicting CTR:
  - Get the product, Pr(examination) x Relevance, right.
- Predicting Relevance:
  - Need to correctly assign credit between examination and relevance.
- Incorrectly assigning credit can improve CTR prediction, while making relevance estimates less accurate.

Predicting relevance.

- Run an experiment on live traffic.
  - Sponsored results are ranked by bid x relevance.
  - More accurate relevance estimates should result in higher CTR and revenue.
    - Will place results with higher relevance in positions with higher Pr(examination).
- Baseline/pure-relevance had better revenue and CTR than max-examination.
  - Results were statistically significant results.

Conclusions

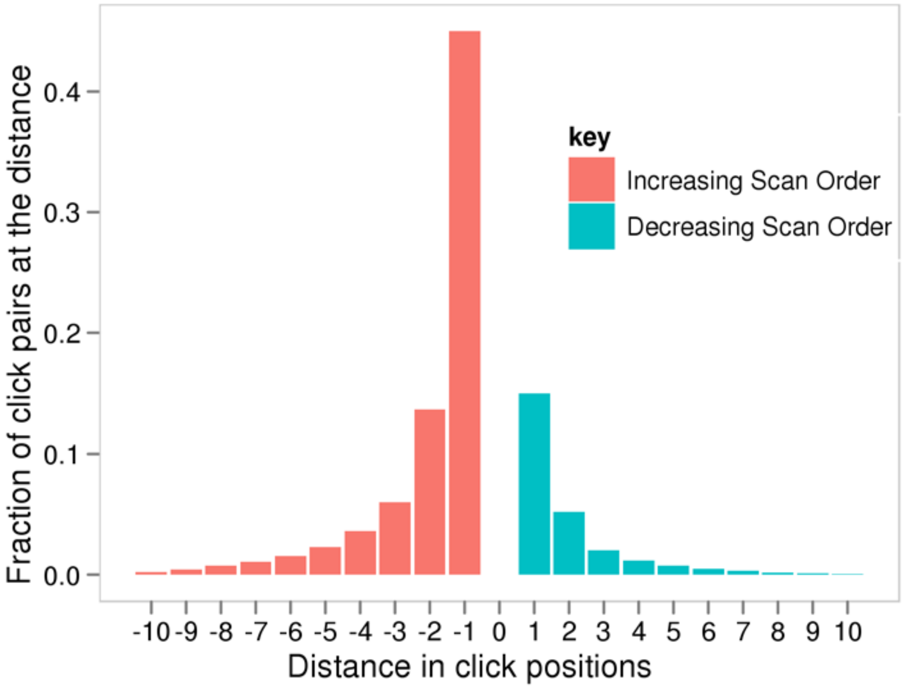
- Changes in CTR when conditioned on other clicks are also due to instance relevance, not just examination.
- New user browsing models that incorporate this insight are more accurate.
- Evaluating user browsing models solely using offline analysis of CTR prediction can be problematic.
  - Use human ratings or live experiments.

Future Work

- What about organic search results?
- Quantitatively assigning credit between instance relevance and examination.
  - Features are correlated.
- Generalize pure-relevance and JRE to incorporate information about the relevance of prior results, or the satisfaction of the user with the prior clicked results.

Backup

Scan order



Fraction of click pairs at the distance
0.0
0.1
0.2
0.3
0.4
key
Increasing Scan Order
Decreasing Scan Order
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Distance in click positions