Improving Ad Relevance in Sponsored Search

Dustin Hillard, Stefan Schroedl, Eren Manavoglu, Hema Raghavan and Chris Leggetter

Motivation

- Improve ad relevance for search users
- Develop a model to predict relevance
- Leverage user interactions in learning
- Use predicted relevance to improve system
  - As a filter to remove bad ads
  - As a feature to improve ad ranking
  - As a score for improving ad page placement

Outline

- Motivation
- Ad Relevance Models
  - Baseline model
  - Learning from user clicks
- Sponsored Search Applications
- Conclusion

Outline

- Motivation
- Ad Relevance Models
  - Baseline model
  - Learning from user clicks
- Sponsored Search Applications
- Conclusion

An Ad Relevance Model

- Develop a model to predict how relevant an ad is for a particular query
- Incorporate typical IR features such as word and character overlap, word novelty
- Train a machine learned model based on human generated editorial judgments

Relevance Modeling Data

- Retrieve about 20 ads per query with a typical information retrieval system
- Stratified query sample from web logs
- Binary ‘good’ vs. ‘bad’ editorial judgments

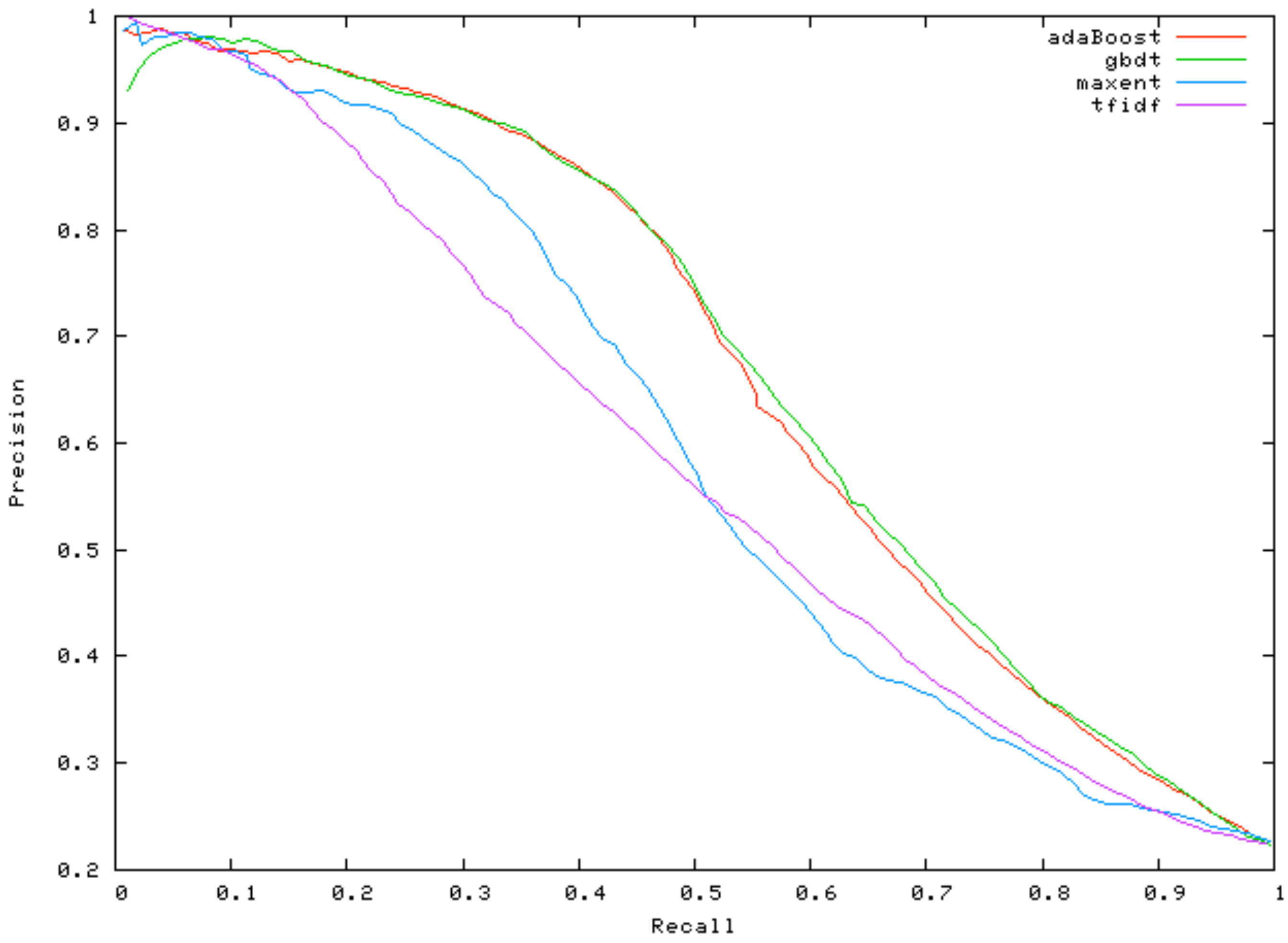
Data	Queries	Query-Ad Pairs
Train	4.8k	95k
Test	2.3k	47k

Baseline Model: Results

- Baseline features:
  - Character, word and bigram overlap
  - Ordered bigram overlap
  - Cosine match (TF/IDF)
  - Query length

	Precision	Recall	F-Score
maxent	0.658	0.458	0.540
adaBoost	0.670	0.543	0.600
GBDT	0.671	0.551	0.605

Baseline Precision/Recall



Outline

- Motivation
- Ad Relevance Models
  – Baseline model
  – Learning from user clicks
- Sponsored Search Applications
- Conclusion

Learning From User Clicks

- Incorporate information from user click behavior to improve relevance modeling
- Include historical click information:
  - Directly for specific observed click rates
  - Broadly with a query->ad click translation model

Observed Click History

- Previous click history is the best predictor of future click behavior
- Collect aggregate click rate statistics from our logs at multiple levels of granularity
  - Query-Ad, Query-Advertiser levels
  - Ad, Advertiser levels
  - Query level
- Broader aggregates are less precise but have higher coverage

Insufficient Click History

- No history is available for previously unseen ads, or infrequent query-ad pairs
- Develop a model that predicts click propensity based only on query-ad text
- Learn a relationship between a query and an ad title that can be applied to unseen query-ad pairs

A Click Translation Model

- Learn a query->title translation model

$$p(D|Q) = p(Q|D)p(D)/p(Q)$$

- IBM Model I, with web logs as corpus

$$p(Q|D) = \prod_{j=0}^{m} \sum_{i=0}^{n} trans(q_j|d_i) \qquad trans(q_j|d_i) = \frac{\sum_{logs} count(q_j|d_i)}{\sum_q \sum_{logs} count(q|d_i)}$$

- Compare 2 models: click-based, view-based

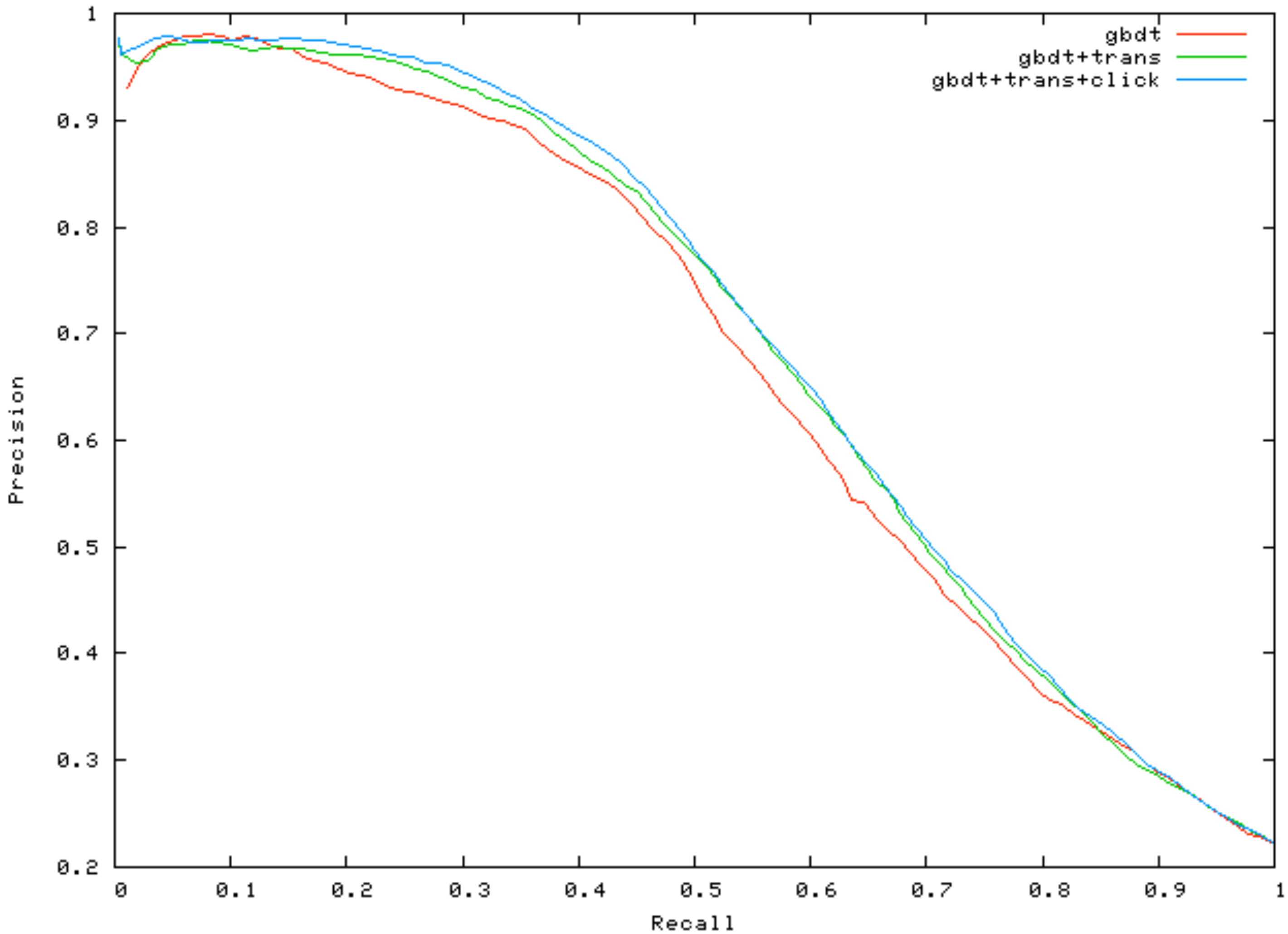
$$clickLikelihood = \frac{p_{click}(Q|D)}{p_{ec}(Q|D)}$$

Using Clicks: Results

- Start with baseline GBDT model
  - Add observed click history features only
  - Add click translation scores only
  - Add both together

	Precision	Recall	F-Score
Baseline (GBDT)	0.671	0.551	0.605
+click history	0.699	0.557	0.620
+translations	0.658	0.590	0.622
+click +trans	0.673	0.584	0.625

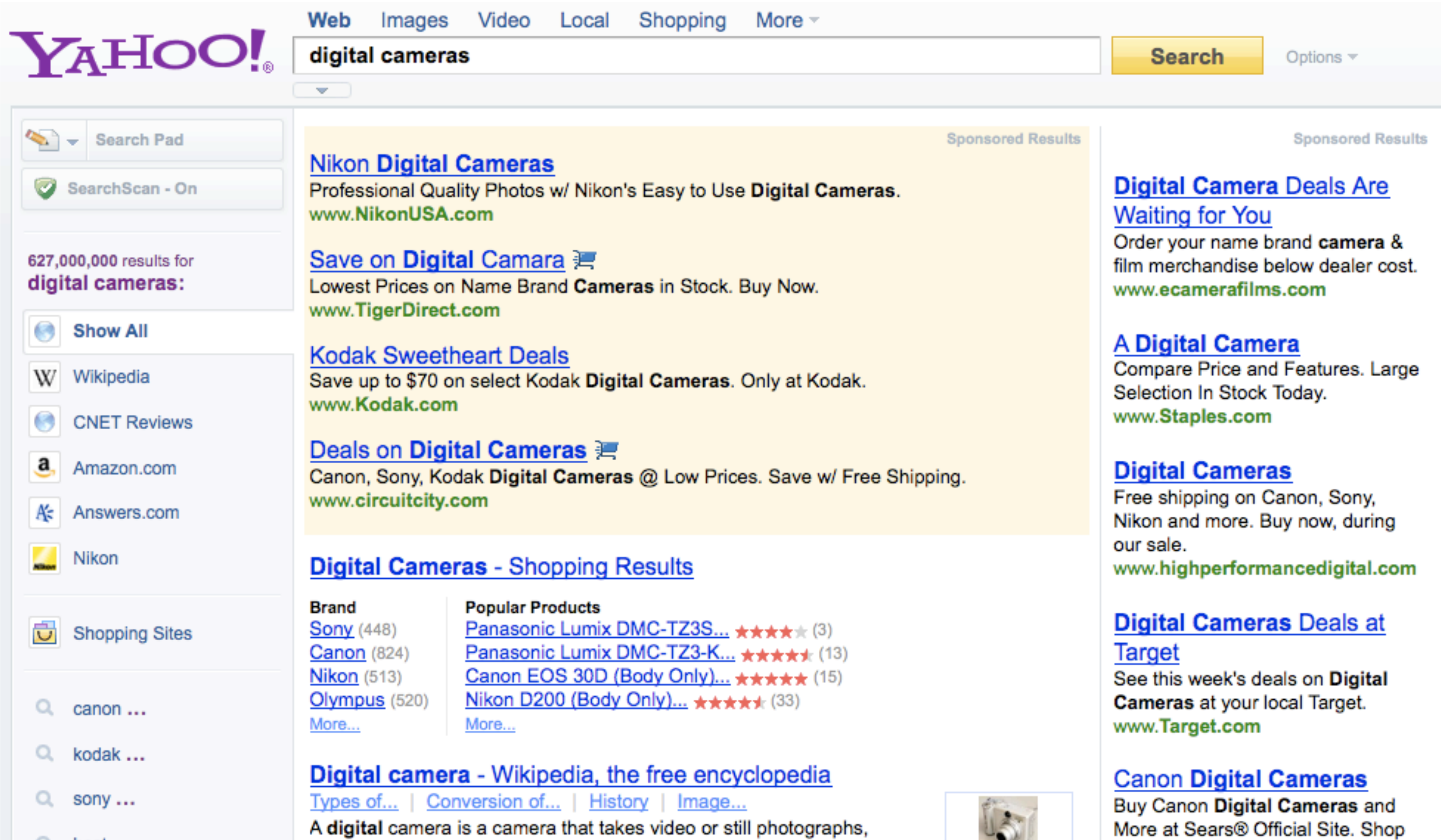
Precision/Recall With Clicks



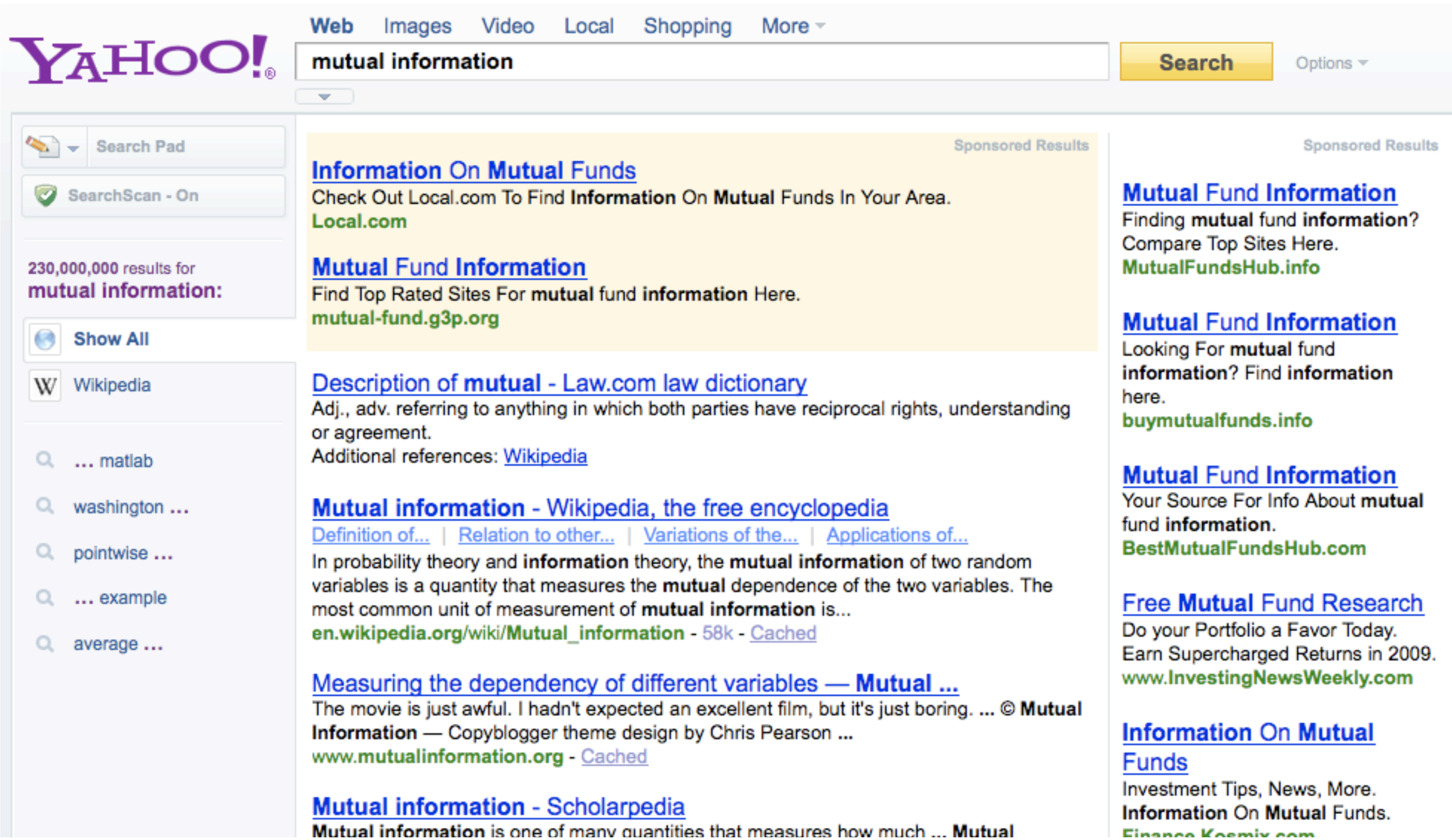
Outline

- Motivation
- Ad Relevance Models
  - Baseline model
  - Learning from user clicks
- Sponsored Search Applications
- Conclusion

Example of Sponsored Search
(typical)



Example of Sponsored Search
(could be better)



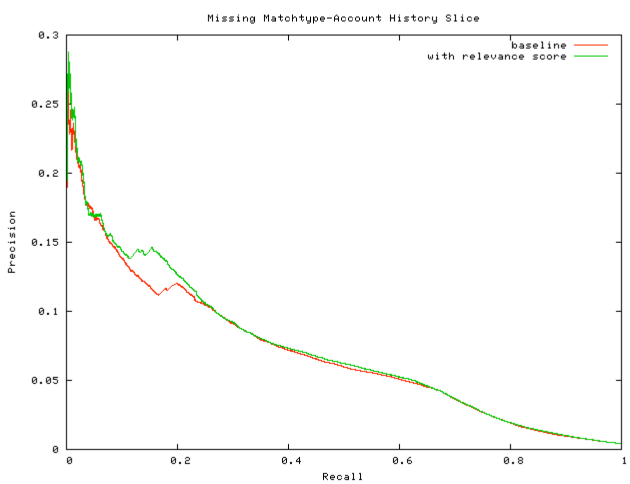
Ad Filtering

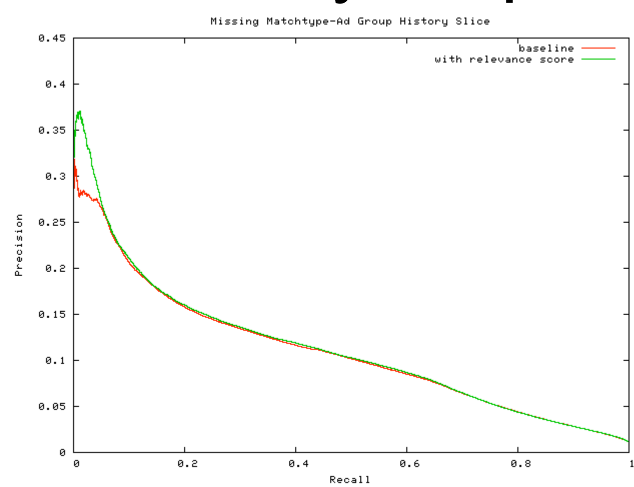
- Problem: Remove low quality ads
- Approach: Filter low relevance score ads
- Impact:
  – Filtered 50% of Bad ads, less than 10% of Good
  – Bucket metrics:

	Relative Change
coverage	-8.7%
ad depth	-11.9%
ad CTR	+10.1%
total ad clicks	+0.5%

Ad Ranking

- Problem: Rank ads by bid and p(click)
- Approach: Provide relevance as feature
- Impact:
  - Improves click model when history is sparse





Optimization

- Problem: Place ads on the search page
- Approach: Consider ad and web relevance
- Impact:
  - Reduced low quality ads above search results
  - Bucket metrics:

	Relative Change
North Ad Impact	-4.5%
North ad CTR	+1.5%
total ad clicks	+0.8%

Outline

- Motivation
- Ad Relevance Models
  - Baseline model
  - Learning from user clicks
- Sponsored Search Applications
- Conclusion

Summary

- Developed a useful ad relevance model
- Improved performance with user click data
- Extend to new ads with click trans. model
- Incorporated in sponsored search system:
  - Removed low quality ads
  - Improved ad ranking
  - Improved ad placement

THANKS!

QUESTIONS?