Computational Advertising

UW CSE454

Thanks To:

Mike Mathieu
mike@frontseat.org

Thanks To:

Dr. Andrei Broder

Dr. Evgeniy Gabrilovich

Dr. Vanja Josifovski

YAHOO! RESEARCH

2012 Global Ad Spend

$530 Billion

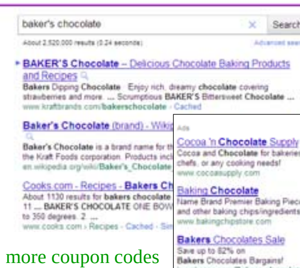
"Half the money
I spend on advertising
is wasted;
**the trouble is I don't know
which half.**"
-- John Wanamaker (attributed)
[1838-1922]

US online advertising spending
(source: eMarketer.com, November 2010)

Year	Online	Online % of total media
2009	$22.7B	13.9%
2010	$25.8B	15.3%
2011	$28.5B	16.7%
2012	$32.6B	18.3%
2013	$36.0B	19.8%
2014	$40.5B	21.5%

 6

What changed in 100 years? measurability and reach



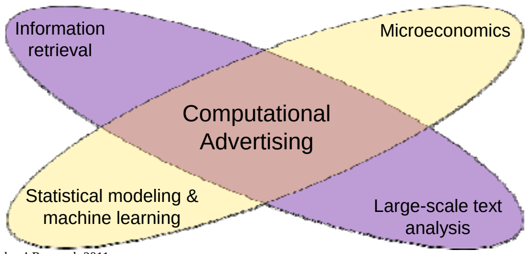
- No more coupon codes
- Flexible ad targeting + conversion tracking
- ***Experimentation rules !***

 7

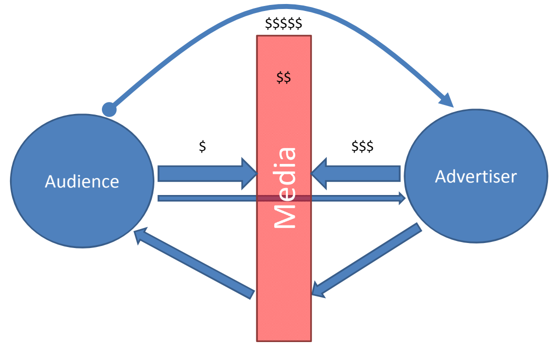
What is "Computational Advertising"?

A new scientific sub-discipline that provides the foundation for building online ad *retrieval* platforms

Find the optimal ad for a given user in a specific context



Industry Structure



The Great Divide

Brand	Direct Response
• Emotions	• Transactions
• Indirect benefits	• Gross profits
• Banners, TV, stadiums	• Search, coupons, 1-800, radio, mail

Anatomy of an ad

Tutorial at **SIGIR 2010**
Information Retrieval Challenges in Computational Advertising
research.yahoo.com/tutorials/sigir

Title
Creative
Display URL
Landing URL

http://research.yahoo.com/sigir10_compadv

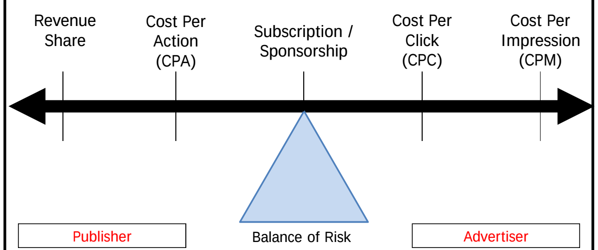
Bid phrases: {SIGIR 2010, computational advertising, Evgeniy Gabrilovich, ...}
Bid: $0.10

So when do advertising dollars actually change hands?

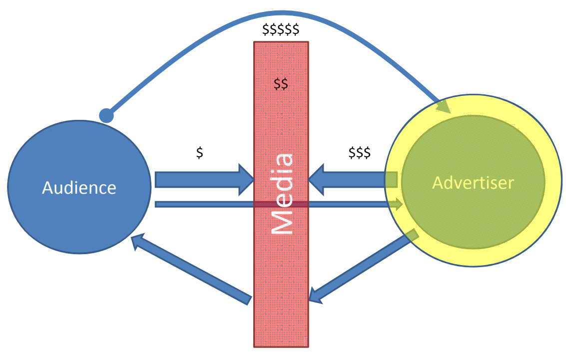
	Risk
CPM = cost per thousand impressions • Typically used for graphical/banner ads (brand advertising)	To the advertiser
CPC = cost per click • Typically used for textual ads	Shared
CPT/CPA = cost per transaction/action • Also known as referral fees or affiliate fees	To the search engine

 14

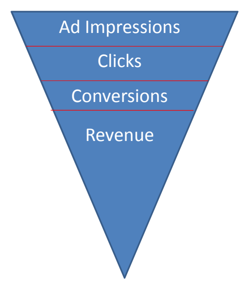
Online Advertising Risks



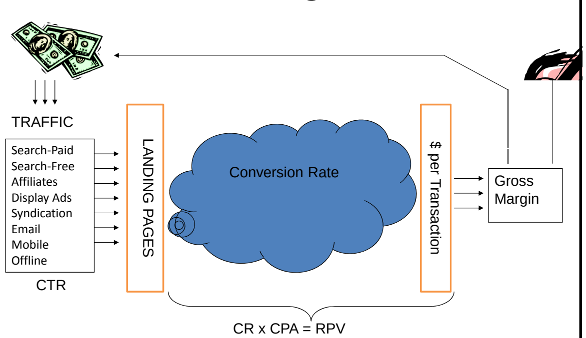
Industry Structure



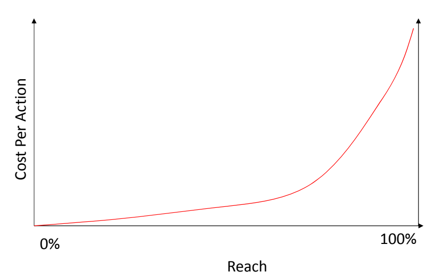
Conversion Funnel



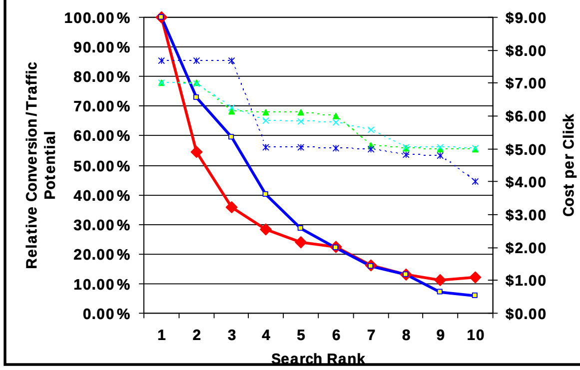
Monetizing Traffic



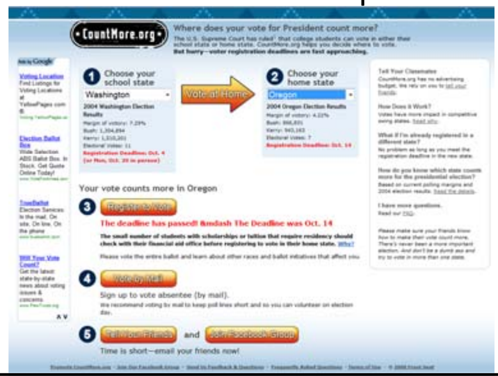
Share of Voice Costs $$$



Conversion Potential vs. Price



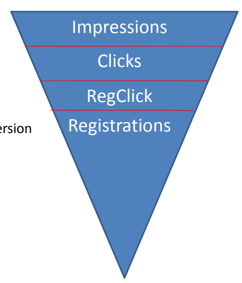
Real World Example



Real World Example

CountMore.org

- RefSrc on URL
- Drop cookie
- Pass RefSrc upon conversion
- Match with ad spend
- Calculate CPA



4.4M
2078
69
29

CTR=0.0469%
CPC=$0.65
eCPM=$0.31
CPRegClick=$19.69
CPReg=$46.76

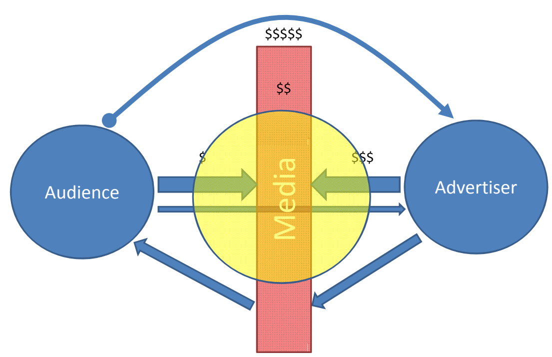
Bid Management

Term	Clicks	CPC	Pos	CR	Leads	CPA	AvgPrice	Revenue	Spend	GM
Nursing School	5,000	$1.00	1	5%	250	$20.00	$7.50	$1,875	$5,000	-63%
Nursing Schools	5,000	$2.00	3	20%	1,000	$10.00	$30.00	$30,000	$10,000	200%
Total	10,000	$1.50	2	12.5%	1,250	$12.00	$25.50	$31,875	$15,000	113%

Bid Management

Term	Clicks	CPC	Pos	CR	Leads	CPA	AvgPrice	Revenue	Spend	GM
Nursing School	5,000	$1.00	1	5%	250	$20.00	$7.50	$1,875	$5,000	-63%
Nursing Schools	5,000	$2.00	3	20%	1,000	$10.00	$30.00	$30,000	$10,000	200%
Total	10,000	$1.50	2	12.5%	1,250	$12.00	$25.50	$31,875	$15,000	113%
Optimized	8,000	$2.43	1	22%	1,760	$11.05	$30.00	$52,800	$19,440	172%

Industry Structure



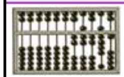
Beyond keyword matching

- Matching ads is relatively simple for explicitly bid keywords
  - **Exact match**
- Covering only those is not enough – advertisers need volume !
  - **Broad match** (or **advanced match**)

- Suppose your ad is "Low prices on Seattle hotels"
- Naïve approach: bid on all queries that contain the word "Seattle"
- **Problems**
  - 'Seattle's Best Coffee Chicago' 
  - 'Alaska cruises start point' 
- **Ideally:** bid on queries related to Seattle as a travel destination
  - The system should facilitate concept-level ad matching

 26

An ultra-brief history of approaches to Web advertising



The old school: database-style ad matching

- Exact match (query = bid phrase)
- Broad match via query rewrites
- Content match: reduce the problem to exact match (extract bid phrases from pages)

➔ **Essentially record lookup**

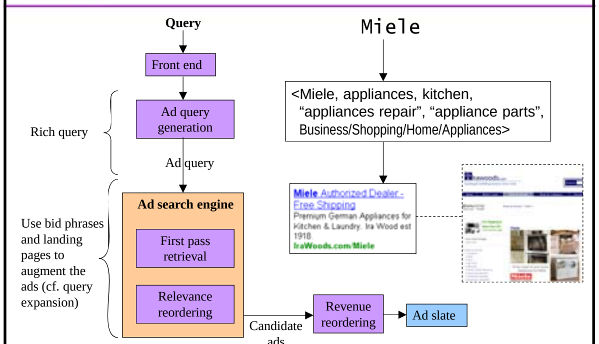
The new approach: knowledge-based ad retrieval

- Elaborate query expansion
- Ad indexing and scoring using all the info available
  - Bid phrases, title, creative, URL, landing page, etc.
  - ➔ Akin to document indexing in IR
- 2nd pass relevance reordering (re-ranking)
  - Using features not available to the 1st pass model (e.g., set-level features, click history)

Learning to rank

 27

Our approach to sponsored search



 28

Ads vs. Web pages

Ads

- Very short
- Optimized for presentation, not for indexing
  - creatives have low SNR
- Legacy bid-phrase-centric definition dictated by the exact match scenario
  - very limiting today
- Complex structure

Web pages

- Not overly short (at least more often than not ☺)
- Simple structure: sections/subsections and (optional) HTML markup

 29

Ad retrieval vs. Web search

Ad retrieval

- Smaller corpus
- Much broader notion of relevance (relatedness)
- Different (but rich) information is available
  - bids, budgets, landing pages, conversion rates, elaborate nested structure of campaigns, …

Web search

- Huge corpus
- Mainly aiming at pages that subsume all the query terms
  - Strict notion of relevance
- Anchor text and other valuable signals are available

 30

RPV Optimization:
Problems with Sort by CPC

Example Term: "mba"		
Ad Title	Univ. of Phx: Online MBA	Univ. of Washington MBA
Ad Body	100% online university. Fully accredited.	Foster School of business. Top 30 ranked.
CPC	$10.00	$0.50

RPV Optimization:
Problems with Sort by CPC

Example Term: "mba"		
Ad Title	Univ. of Phx: Online MBA	Univ. of Washington MBA
Ad Body	100% online university. Fully accredited.	Foster School of business. Top 30 ranked.
CPC	$10.00	$0.50
CTR	0.01%	4%
Position	#1	#10
RPV	$0.0010	$0.0200

Should we show ads at all

Learning when (not) to advertise
(CIKM 2008, Broder et al.)

- **One does not *have* to show ads!**
  - Roughly **half** of the queries have no ads

- Repeatedly showing non-relevant ads can have detrimental long-term effects
  - Modeling actual (short- and long-term) costs of showing non-relevant ads is very difficult
- **Goal: predict when (not) to show ads**

Should we show ads at all

Two approaches: Thresholding vs. Machine Learning

- **Global threshold** on relevance scores of **individual ads**
  - Only show ads with scores above the threshold
- **Problem:** Scores are not necessarily comparable across queries!

- **Learn** a binary prediction model for **sets of ads**
- Features defined over **sets of ads** rather than individual ads
  - **Relevance** (word overlap, cosine similarity between ad and query/page etc.)
  - **Result set cohesiveness** (coefficient of variation of ad scores, result set clarity, entropy)

Features

- Relevance features
  - Word overlap, cosine similarity between ad and query/page
- Vocabulary mismatch features
  - Translation models
  - PMI between query/page terms and bid terms
- Ad-based features
  - Bid price (higher bids often indicate better ads)
- Result-set cohesiveness features
  - Coefficient of variation of ad scores (std/mean)
  - Result set clarity
    - If the set of ads is very cohesive and focused on 1-2 topics, the relevance language model is very different from the collection model
  - Entropy

Should we show ads at all

Incorporating click history
(WSDM 2010, Hillard et al.)

- Binary classifier (relevant / non-relevant ads)
  - Baseline: text overlap features (query/ad)
- Click history (query/ad) with back-off
- Click propensity in query/ad translation

Bayes' rule

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

IBM Model 1

$$p(Q|D) = \prod_{j=0}^{m}\sum_{i=0}^{n} trans(q_j|d_i)$$

$$trans(q_j|d_i) = \frac{\sum_{\text{logs}} count(q_j|d_i)}{\sum_q \sum_{\text{logs}} count(q_j|d_i)}$$

Counting clicks for query/ad word pairs

- Cold start (i.e., no click history) is OK
- Using click data to overcome synonymy
  Query = "running gear"
  Ad = "Best jogging shoes"

Results

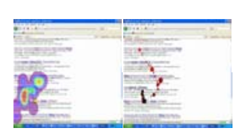
Query coverage	↓	**9%**
Ads per query	↓	**12%**
CTR	↑	**10%**

Same # clicks on fewer ads

Should we show ads at all

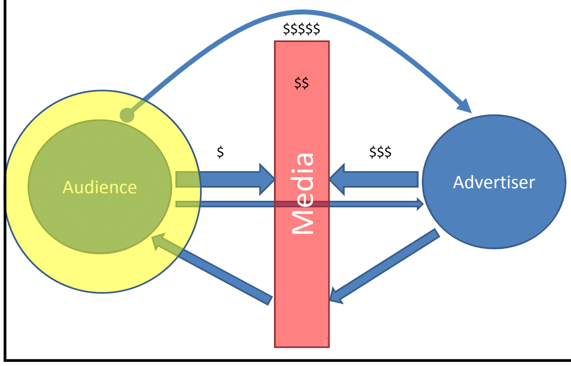
Incorporating multi-modal interaction data
(SIGIR 2010, Guo & Agichtein)

- Ready to buy or just browsing ?
  - Classifying research- and purchase-oriented sessions
- **Inferring** eye gaze position from **observable** actions
  - Keystrokes, GUI (scroll/click), mouse movement, browser (new tab, close, back/forward)



- Research vs. purchase classification (in lab): F1 = 0.96
- Ad clickthrough in sessions classified as *Purchase* > 2X compared to sessions classified as *Research*
- Predicting future ad clicks: F1 = 0.07 ↗ 0.17 (**+141%**)

Industry Structure



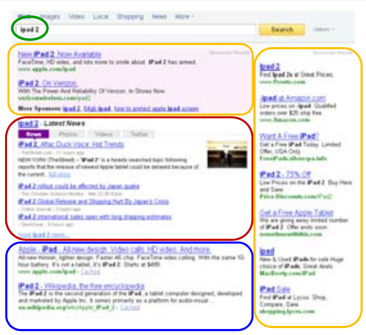
End Users

Don’t bug me

Unless I like what you have to offer

Ads as another source of content for enriching Web search results

“I do not regard advertising as entertainment or an art form, but as a medium of information….”



 15

Better Matching

- Context detection
  - GPS, location
  - App vs. content
  - In-game
  - Info seeker vs. transactor
  - Calendars/schedules/events
  - Social networks/status
  - Twitter - now
  - Behavioral – esp. w/knowledge of specific site behaviors
  - Contextual
- Privacy
  - Google “AOL search data”

Textual advertising

Sponsored Search

Ads driven by search keywords
a.k.a. *“keyword driven ads” or “paid search”*

Content Match

Ads driven by the content of a web page
a.k.a. “*context driven ads” or “contextual ads”*

Textual advertising on the Web is strongly related to NLP and information retrieval

 47

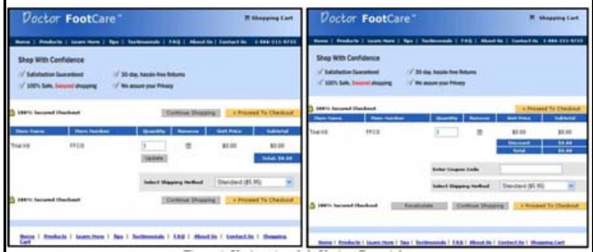
Context?

- Flowers
- Mentos gum
- Trial Prep
- Credit score
- Cosmetics
- Hampton Inns
- WeightWatchers
- Vacation Home Rentals
- Home Depot
- Web Hosting
- WebMD
- Colon Cleanse – Warning
- My Teeth Aren’t Yellow
- Classmates.com

TESTING

One accurate measurement is worth more than a thousand expert opinions
— Admiral Grace Hopper

Nine Differences



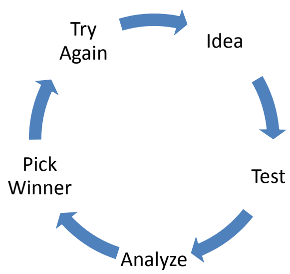
A

upgrade!

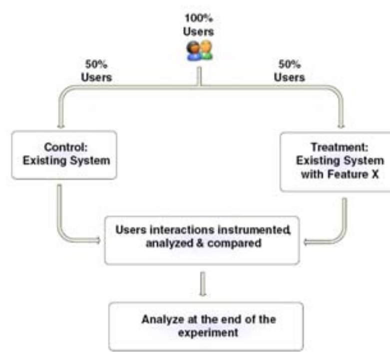
B

Lost 90% revenue….
Reverting coupon code increased CR 6.5%

Testing



A/B Split Test



Testing

Sample Size, margin of error, confidence

$x = Z(^{c}/_{100})^2 r(100\text{-}r)$

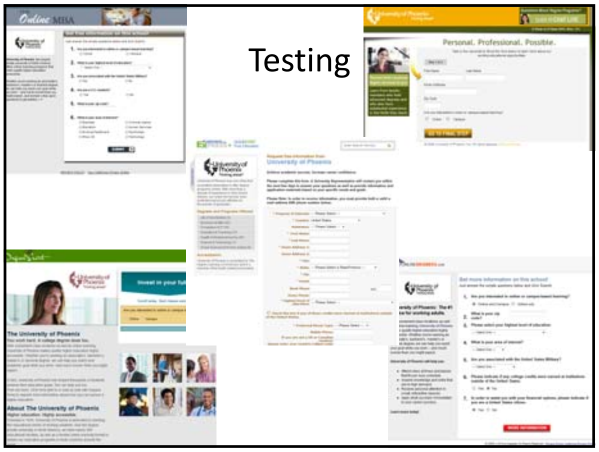
$n = {}^{N\,x}/_{((N\text{-}1)E^2 + x)}$

$E = \text{Sqrt}[^{(N\,-\,n)x}/_{n(N\text{-}1)}]$

Sample Size Problems

- So many ideas, so little to sample…
  - Disproportionate advantage to scale
- Multivariate testing
  - Taguchi Method
    - Method for calculating signal-to-noise ratio of different parameters in an experimental design
    - Allows optimization with A/B test of each cross-product

Testing



Repetition

Professional Photos

Before

We observed an immediate 30% increase in conversion rates

After

Fact Sheet Design

Existing Schools (n=1,428)	CR
Best	51.1%
Worst	0.4%
Average	11.6%

Test	# Schools	CR Lift
Professional photo	1	30%
More RFI buttons	3	21%
Marketing voice, more programs listed	1	28%
Photos + Marketing voice, more programs	1	50%

Google Analytics

- Why

Opportunities Today

- Conversions
  - Low-RPV
  - Waste
  - Simplicity
- Risk
  - Scaling local, hyperlocal
  - Data exchanges
  - Under-monetized sites
- Context

Summary

- Conversions
- Risk
- Context
- Testing