Computational Advertising

UW CSE454

Thanks To:

Mike Mathieu
mike@frontseat.org

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)

<table>
<thead>
<tr>
<th>Year</th>
<th>Online</th>
<th>Online % of total media</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>$22.7B</td>
<td>13.9%</td>
</tr>
<tr>
<td>2010</td>
<td>$25.8B</td>
<td>15.3%</td>
</tr>
<tr>
<td>2011</td>
<td>$28.5B</td>
<td>16.7%</td>
</tr>
<tr>
<td>2012</td>
<td>$32.6B</td>
<td>18.3%</td>
</tr>
<tr>
<td>2013</td>
<td>$36.0B</td>
<td>19.8%</td>
</tr>
<tr>
<td>2014</td>
<td>$40.5B</td>
<td>21.5%</td>
</tr>
</tbody>
</table>

© Yahoo! Research 2011

What changed in 100 years?
measurability and reach

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

Statistical modeling & machine learning
Large-scale text analysis
Information retrieval
Microeconomics
Computational Advertising

The Great Divide

Brand
- Emotions
- Indirect benefits
- Banners, TV, stadiums

Direct Response
- Transactions
- Gross profits
- Search, coupons, 1-800, radio, mail

Industry Structure

Audience
Media
Advertiser

The New York Times

Anatomy of an ad

Title
Creative
Display URL
Landing URL

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

© Yahoo! Research 2011
So when do advertising dollars actually change hands?

**CPM = cost per thousand impressions**
- Typically used for graphical/banner ads (brand advertising)

**CPC = cost per click**
- Typically used for textual ads

**CPT/CPA = cost per transaction/action**
- Also known as referral fees or affiliate fees

---

### Online Advertising Risks

<table>
<thead>
<tr>
<th>Revenue Share</th>
<th>Cost Per Action (CPA)</th>
<th>Subscription / Sponsorship</th>
<th>Cost Per Click (CPC)</th>
<th>Cost Per Impression (CPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>Balance of Risk</td>
<td>Advertiser</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Industry Structure

- **Audience**
- **Advertiser**
- **Media**

---

### Conversion Funnel

- Ad Impressions
- Clicks
- Conversions
- Revenue

---

### Monetizing Traffic

- **Traffic**
- Search Paid
- Search Free
- Affiliates
- Display Ads
- Syndication
- Email
- Mobile
- Offline

- **CTR**
- **Conversion Rate**
- **Gross Margin**

- CR x CPA = RPV

---

### Share of Voice Costs $$$

- Cost Per Action
- Reach
- 100%
- 0%
**Conversion Potential vs. Price**

- 0% to 100% Conversion Potential vs. $0.00 to $9.00 Cost per Click.

**Real World Example**

- Impressions: 4.4M
- Clicks: 2,078
- RegClick: 69
- Registrations: 29

CTR=0.0469%
CPC=$0.65
eCPM=$0.31
CPCRegClick=$19.69
CPRg=$4.76

**Bid Management**

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

**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
    - The system should facilitate concept-level ad matching

Our approach to sponsored search

- Query
  - From end
  - Rich query
    - Use bid phrases and landing pages to augment the ad (cf. query expansion)
  - Ad query generation
  - Ad search engine
  - First pass retrieval
  - Relevance modeling
- Miele
  - <Miele, appliances, kitchen, 
    “appliances repair”, “appliance parts”, Business/Shopping/Home/Appliances>

Ads vs. Web pages

- Ads
  - Very short
  - Optimized for presentation, not for indexing
  - Creative ads 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
  - Anchor text and other valuable signals are available

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

RPV Optimization: Problems with Sort by CPC

<table>
<thead>
<tr>
<th>Ad Title</th>
<th>Univ. of Phx: Online MBA</th>
<th>Univ. of Washington MBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Body</td>
<td>100% online university. Fully accredited.</td>
<td>Foster School of Business. Top 30 ranked.</td>
</tr>
<tr>
<td>CPC</td>
<td>$10.00</td>
<td>$0.50</td>
</tr>
</tbody>
</table>

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
  - 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

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)
RPV Optimization: Problems with Sort by CPC

Example Term: "mba"

<table>
<thead>
<tr>
<th>Ad Title</th>
<th>Univ. of Phx. Online MBA</th>
<th>Univ. of Washington MBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad Body</td>
<td>100% online university. Fully accredited.</td>
<td>Foster School of business. For 30 ranked.</td>
</tr>
<tr>
<td>CPC</td>
<td>$10.00</td>
<td>$0.50</td>
</tr>
<tr>
<td>CTR</td>
<td>0.01%</td>
<td>4%</td>
</tr>
<tr>
<td>Position</td>
<td>#1</td>
<td>#10</td>
</tr>
<tr>
<td>RPV</td>
<td>$0.0010</td>
<td>$0.0200</td>
</tr>
</tbody>
</table>

• 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

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

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

\[
\frac{p(\text{click} | \text{query})}{p(\text{click} | \text{ad})} = \frac{p(\text{query})}{p(\text{ad})} \cdot \frac{p(\text{click} | \text{query})}{p(\text{click} | \text{ad})} 
\]

Bayes’ rule

IBM Model I

\[
\text{trans(q, d)} = \frac{\text{count}(q, d)}{\sum_d \text{count}(q, d)} 
\]

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 \(\pm 0.17 \ (\pm141\%)

Should we show ads at all

IBM Model I

\[
\text{trans(q, d)} = \frac{\text{count}(q, d)}{\sum_d \text{count}(q, d)} 
\]

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
2/19/2013

Industry Structure

Audience $ $ $$ $ $$$ $ $$ 

Media $ $$$ 

Advertiser $ $$$ 

End Users

Don’t bug me

Unless I like what you have to offer

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
Testing

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

Nine Differences

A
B

upgrade!

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

Testing

Try Again

Idea

Pick Winner

Test

Analyze

A/B Split Test

50% Users

100% Users

Control: Existing System

Treatment: Existing System with Feature X

Users interactions instrumented, analyzed & compared

Analyze at the end of the experiment

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

Sample Size, margin of error, confidence

\[ n = \frac{N \cdot \sigma^2}{\left( N - 1 \right) E^2 + x} \]

\[ \sigma = \frac{Z\left( \frac{\alpha}{2} \right)}{\sqrt{2}} \]

\[ E = \sqrt{\frac{N \cdot x}{n(N-1)}} \]
Testing

Repetition

Professional Photos

Fact Sheet Design

Google Analytics

<table>
<thead>
<tr>
<th>Test</th>
<th># Schools</th>
<th>CR Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional photo</td>
<td>1</td>
<td>30%</td>
</tr>
<tr>
<td>More RF1 buttons</td>
<td>3</td>
<td>21%</td>
</tr>
<tr>
<td>Marketing voice, more programs listed</td>
<td>1</td>
<td>28%</td>
</tr>
<tr>
<td>Photos + Marketing voice, more programs</td>
<td>1</td>
<td>50%</td>
</tr>
</tbody>
</table>

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

We observed an immediate 30% increase in conversion rates
<table>
<thead>
<tr>
<th>Opportunities Today</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Conversions</td>
<td>• Conversions</td>
</tr>
<tr>
<td>– Low-RPV</td>
<td>• Risk</td>
</tr>
<tr>
<td>– Waste</td>
<td>• Context</td>
</tr>
<tr>
<td>– Simplicity</td>
<td>• Testing</td>
</tr>
<tr>
<td>• Risk</td>
<td></td>
</tr>
<tr>
<td>– Scaling local, hyperlocal</td>
<td></td>
</tr>
<tr>
<td>– Data exchanges</td>
<td></td>
</tr>
<tr>
<td>– Under-monetized sites</td>
<td></td>
</tr>
<tr>
<td>• Context</td>
<td></td>
</tr>
</tbody>
</table>