Outline

- “Science will replace much of the art of marketing” – Eric Schmidt
- Introduction
  - Online Advertising and SEM basics
- Search Engine Perspective in SEM
  - Mechanism Design and Auction Rules
  - Challenges
  - Click Fraud
- Advertiser Perspective
  - Keyphrase Generation
  - Bidding
- Sponsored Search with Context
- Empirical Issues in Sponsored Search
- Conclusions
Introduction

Traditional Advertising

- Print, TV, Radio, Direct Mail
- Ad agencies buy media on behalf of firms
  - Dominated by large holding companies (Omnicom, WPP, InterPublic, Publicis Groupe)
  - Individual agencies like JWT, Ogilvy & Mather, McCann Erickson manage branding
  - Low barriers to entry (talent and sales)
    » But some economies of scale
- Local Advertising (YPs)
  - Publisher centric
Traditional Advertising

- Pricing
  - Traditionally fixed commissions
  - Later negotiated commissions
  - Recently, labor-based + performance-based
    - Pushed by advertisers: aligns incentives
    - Reflects growth in promotions, event marketing, product placement

Advertising spend

<table>
<thead>
<tr>
<th>Media</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers</td>
<td>47</td>
</tr>
<tr>
<td>Direct Mail</td>
<td>56</td>
</tr>
<tr>
<td>Broadcast TV</td>
<td>45</td>
</tr>
<tr>
<td>Radio</td>
<td>20</td>
</tr>
<tr>
<td>Cable TV</td>
<td>22</td>
</tr>
<tr>
<td>Magazine</td>
<td>16</td>
</tr>
<tr>
<td>Yellow Pages</td>
<td>15</td>
</tr>
<tr>
<td>Internet</td>
<td>17</td>
</tr>
<tr>
<td>TOTAL</td>
<td>238</td>
</tr>
</tbody>
</table>
Online advertising

- Banner ads (Doubleclick)
  - Standardized ad shapes with images
  - Loosely related to content
- Search linked ads (Google Adwords)
  - Related to search terms
- Context linked ads (Google AdSense)
  - Related to content on page
  - Closely tied to Google’s existing competence on figuring out what a page is really about

Search Engine Marketing (SEM)
Search Engine Marketing

- High relevance (based on user context)
- Occurs close to decision/buying time
- Highly measurable
- Results in very good performance in cost-per-acquisition (adv expenditure per sale)
- Rapid growth but still small part of advertising market
  - Total US adv spend ~ $220-240 billion, growth ~ 1-2% per year
  - Online advertising: ~ $17 billion, growth ~ 35% last year

Google Valuation

[Graph showing Google valuation over time]
Information Retrieval (IR)

- Crawler machines
  - crawl the web
  - Check for duplicates, store the documents

- Search engine servers
  - Create an inverted index
  - Inverted index
  - Docs

- User query
  - Show results to user

Source: Marti Hearst (SIMS, UC Berkeley)

Share of Searches

- Organic Traffic is a key to driving SEM revenues

Source: Search Engine Watch (2006)
How does SEM work?

- **Search engines**
  - run keyword auctions to sell available inventory of ad positions

- **Advertisers**
  - submit bids which indicate their willingness-to-pay per click
    - for example, bid of $2.10 per click for the keyword "laptop"

- The search engine orders the ads in descending order
  - Bid is a key determinant of ad position
  - Other factors such as CTR are also factored in
    - More on the mechanism later

Search Engine Perspective
The Advertiser Perspective

Advertiser Perspective

- Choose keywords
- Set budget
  - Budgets are commonly observed
  - Proxies for inventory constraints, short-term capital constraints, etc
  - Sometimes, to protect against click-fraud
- Determine bid (maxCPC) for each keyword
  - Auctions not incentive compatible
  - Budget constraints may restrict bidding strategies
- Role of SEM firms
Keyphrase Generation

Current Methods

- Query log and Advertiser log mining
  - Mine query logs and logs of advertiser searches
    - Google and Yahoo
  - Mine other keyphrases advertisers have bid on
    - Yahoo
- Proximity search
- Meta-tag spidering
  - Extract meta-tags from top web pages related to seed keyword
  - Word tracker
- Terms Net (Joshi and Motwani 2006)
  - Based on semantic similarity between keyphrases
Semantic Techniques for Keyphrase Generation

An automated system for relevant keyword generation.

Three Steps:
- Dictionary Creation
- Semantic Similarity
- Keyword Suggestion

Dictionary Creation
Semantic Similarity

\[ K(x, y) = QE(x)QE(y) \]

Search the web for the term \( x \)

Create L2 Normalized vector for \( x \) - \( QE(x) \)

Create L2 Normalized vector for \( y \) - \( QE(y) \)

Compute Similarity using cosine measure

Keyword Generation

- Construct similarity graph using the association matrix.
  - Edge weight inversely related to similarity score
- Traverse the graph using the watershed algorithm.
- Assumption – words with lower frequency are cheaper.
Examples

**skin**

<table>
<thead>
<tr>
<th>skincare</th>
<th>body</th>
<th>aromatherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>facial</td>
<td>scalp skin</td>
<td>hair removal</td>
</tr>
<tr>
<td>treatment</td>
<td>skin smooth</td>
<td>massage</td>
</tr>
<tr>
<td>face</td>
<td>scalp skin</td>
<td>healthcare</td>
</tr>
<tr>
<td>care</td>
<td>skin smooth</td>
<td>microdermabrasion</td>
</tr>
<tr>
<td>occitane</td>
<td>acid peels</td>
<td>natural</td>
</tr>
<tr>
<td>exfoliator</td>
<td>free cosmetics</td>
<td>permanent makeup</td>
</tr>
<tr>
<td>dermal</td>
<td>makeup</td>
<td>aging</td>
</tr>
<tr>
<td>wrinkles</td>
<td>conditioner</td>
<td>shaving</td>
</tr>
<tr>
<td>dermal</td>
<td>wrinkles</td>
<td>treatment glycolic</td>
</tr>
</tbody>
</table>
Budget Constrained Bidding

A Toy Problem

- The advertiser has a daily budget of $1000
- She is considering bidding on four keywords
  - Laptop
  - Refurbished Laptop
  - HP Laptop
  - Sony Laptop
- Assume the bids submitted by competitors are known to the advertiser
- How should the advertiser bid to get the maximum number of clicks (visitors from the search engine ads) each day?
- In real life, 1,000s of keywords and millions of $ per month…
Bid Landscape ($ per click)

- Advertiser's view of other advertisers' bids for the same keywords
- Found via query of Yahoo or Google's advertising bid database

<table>
<thead>
<tr>
<th>Slot</th>
<th>Laptop</th>
<th>Refurbished Laptop</th>
<th>HP Laptop</th>
<th>Sony Laptop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.73</td>
<td>1.29</td>
<td>0.59</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>6.72</td>
<td>0.95</td>
<td>0.58</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>2.00</td>
<td>0.80</td>
<td>0.56</td>
<td>0.84</td>
</tr>
<tr>
<td>4</td>
<td>1.70</td>
<td>0.79</td>
<td>0.55</td>
<td>0.70</td>
</tr>
<tr>
<td>5</td>
<td>1.56</td>
<td>0.78</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>6</td>
<td>1.42</td>
<td>0.77</td>
<td>0.42</td>
<td>0.46</td>
</tr>
<tr>
<td>7</td>
<td>1.13</td>
<td>0.57</td>
<td>0.40</td>
<td>0.43</td>
</tr>
<tr>
<td>8</td>
<td>0.95</td>
<td>0.54</td>
<td>0.40</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Expected number of clicks for each keyword and slot

- Advertiser's view of the benefit of obtaining a given slot for a keyword
- Found via query of Yahoo or Google's advertising bid database
- And through advertiser's own historical clickthrough data

<table>
<thead>
<tr>
<th>Slot</th>
<th>Laptop</th>
<th>Refurbished Laptop</th>
<th>HP Laptop</th>
<th>Sony Laptop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1539</td>
<td>370</td>
<td>45</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>1099</td>
<td>264</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>785</td>
<td>189</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>561</td>
<td>135</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>401</td>
<td>96</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>286</td>
<td>69</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>204</td>
<td>49</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>146</td>
<td>35</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Tradeoffs to make when deciding what to bid?

<table>
<thead>
<tr>
<th>Slot</th>
<th>Laptop</th>
<th>Refurbished Laptop</th>
<th>HP Laptop</th>
<th>Sony Laptop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bid</td>
<td>Clicks</td>
<td>Bid</td>
<td>Clicks</td>
</tr>
<tr>
<td>1</td>
<td>6.73</td>
<td>1539</td>
<td>1.29</td>
<td>370</td>
</tr>
<tr>
<td>2</td>
<td>6.72</td>
<td>1099</td>
<td>0.95</td>
<td>264</td>
</tr>
<tr>
<td>3</td>
<td>2.00</td>
<td>785</td>
<td>0.80</td>
<td>189</td>
</tr>
<tr>
<td>4</td>
<td>1.70</td>
<td>561</td>
<td>0.79</td>
<td>135</td>
</tr>
<tr>
<td>5</td>
<td>1.56</td>
<td>401</td>
<td>0.78</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>1.42</td>
<td>286</td>
<td>0.77</td>
<td>69</td>
</tr>
<tr>
<td>7</td>
<td>1.13</td>
<td>204</td>
<td>0.57</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>0.95</td>
<td>146</td>
<td>0.54</td>
<td>35</td>
</tr>
</tbody>
</table>

Problem formulation

- **Decision variables**
  - how much to bid on each of four keywords

- **Objective function**
  - maximize total number of clicks

- **Constraints**
  - budget – maximum $1,000 spent
  - assignment constraints
    - at most one bid for each key word
Decision variables and objective for the problem

- **Decision Variables**

  \[ x_{ij} = 1 \text{ if slot } i \text{ is assigned to keyword } j, \text{ 0 otherwise;} \]

  where \( i = 1, 2, ..., 8 \) represents the eight slots

  and \( j = A, B, C, D \) represents the 4 keywords.

- **Objective Function**

  \[
  \max 1539x_{1A} + 370x_{1B} + 45x_{1C} + 38x_{1D} + 1099x_{2A} + \ldots + 4x_{8C} + 4x_{8D}
  \]

Constraints for the IP

- **Each keyword must be assigned to one, and only one slot:**

  \[
  x_{1A} + x_{2A} + x_{3A} + \ldots + x_{8A} \leq 1 \quad \text{(Keyword A)}
  \]

  \[
  x_{1B} + x_{2B} + x_{3B} + \ldots + x_{8B} \leq 1 \quad \text{(Keyword B)}
  \]

  \[
  \vdots
  \]

  \[
  x_{1D} + x_{2D} + x_{3D} + \ldots + x_{8D} \leq 1 \quad \text{(Keyword J)}
  \]

- **Total Budget for the day = $1000:**

  \[
  x_{1A}b_{1A} + x_{1B}b_{1B} + \ldots + x_{8D}b_{8D} \leq 1000
  \]

  where \( b_{ij} \) is 1 cent more than the bid of the advertiser currently ranked \( i \) for keyword \( j \)

- **Integrality**

  \( x_{ij} \) are all either 0 or 1
Issues with Formulation?

- Bid landscape typically unknown
- Clicks Vs Position unknown
- Keywords are not of same quality
  - Conversions matter
- IP not scalable

Approach 2: Knapsack

- Knapsack problem and greedy heuristic
- Adaptive algorithm for keyword selection (Paat R* & Williamson)
  - Static model: case where the CTR for the keywords is known
    - Sort keywords in descending order of profit/cost.
    - Choose keywords in descending order (of profit/cost) until the expected cost is close to the budget.
  - Adaptive algorithm works when CTR is not known apriori
    - Use data gathered in the first few periods to estimate the CTR

- Issues: Does not account for slots
Other Techniques for Bidding

- Online knapsack Problem (Chakrabarty et al. 2007)
- Stochastic bidding (Muthukrishnan et al. 2007)
- Efficient Frontier (Holthausen and Assmus 1982)
- Convex Programming: Scalable and accounts for slots (Hosanagar and Stavrinides 2006)

Role of Machine Learning

- "Standard" Optimization techniques solve deterministic problems
- Even stochastic optimization techniques generally assume parameters can be estimated
- Parameters
  - Bid-Rank
  - Rank-Clicks
  - Clicks-Conversions
  - Bid-CPC
Role of Machine Learning

**Metrics:**

<table>
<thead>
<tr>
<th>Impressions</th>
<th>Clicks</th>
<th>Conversions</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searches</td>
<td>Clicks (Bid)</td>
<td>Conversions</td>
<td>Cost</td>
</tr>
<tr>
<td>Keyword</td>
<td>Impressions</td>
<td>Click</td>
<td>Click</td>
</tr>
<tr>
<td>Inventory</td>
<td>CTR</td>
<td>Yield</td>
<td>CPC</td>
</tr>
</tbody>
</table>

**Real-world Complexities**

- “Exact match”
  - Competing with other advertisers you do not know of
- “Broad Match”
  - Your rank changes based on the search query
- Data Integrity
  - Impressions < clicks
  - CPC > Bid
- Data Sparseness
Understanding the Role of SEM Firms

- SEM firms use a variety of techniques to
  - Identify keywords to bid on
  - Bid for each keyword
  - Provide metrics

- Unable to easily generate business online
- Search auctions are too complex for many advertisers
- Bidding Blindly on Search Engines
- RoI hard to measure
- Publishers User-centric, not business-centric

SEM Firms’ Value Proposition

Aggregating the fragmented search channels

Online Users
Generating Keywords

Keywords Created
- cholesterol
- heart medication
- blood pressure medication
- aspirin
- diabetes
- diet
- obesity
- electric help in philadelphia

Determining Bids

<table>
<thead>
<tr>
<th>Slot</th>
<th>Cholesterol</th>
<th>Blood Pressure</th>
<th>Lipid Reduction</th>
<th>Heart Medication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bid</td>
<td>6.73</td>
<td>1.29</td>
<td>0.59</td>
<td>0.93</td>
</tr>
<tr>
<td>Expected Position</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Measuring Conversions

**Conversion** > Phone calls, emails or conversion pages (forms)

Key Metrics

- **Key Levers**
- **Organic traffic:**
  - Metrics: monitor number of pages indexed (benchmark against competitors and look at trends), pagerank (changes slowly), weekly/monthly rank on search engines for your keyword universe (avg, weighted avg pos)
- **Paid traffic?**
Key Metrics

**Metrics:**

<table>
<thead>
<tr>
<th>Impressions</th>
<th>Clicks</th>
<th>Conversions</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keyword</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impressions</td>
<td></td>
<td>Click</td>
<td>Click</td>
</tr>
<tr>
<td>CTR</td>
<td></td>
<td>Yield</td>
<td>CPC</td>
</tr>
</tbody>
</table>

- **Quantity**
- **Quality**
- **Cost**

**Key Metrics**

- **Cost Per Click (CPC) or Per Conversion by Search Engine**
  - These should be similar

- **Conversion Quality by Search Engines**
  - Determines bid intensity

- **Click/conversion Volume Generated**
  - Break down by source

- **Average Position (Rank) by Category**
  - Align with marketing goals
    - Branding (firm level)
    - Promoting specific prod lines
    - LTV of customer
    - Revenues
    - Margins
Sponsored Search with Contexts

- Context: “any auxiliary information that might accompany a search, and might include information that is factual, estimated or inferred. Natural examples of contexts include the zip code, gender, or abstract “intentions”…” (Even-dar et al. 2007)
- Real-world examples
  - Bid on “dentist” but have different bids for different regions (identified by IP addresses)
  - Retailer bids on clothing but places higher premium on women under 30
Context-based Auctions: Market Opportunity

**Local Advertising Market**

$100 B

220 M

Consumer Searches (in MM)

2004 2005 2006 2007 2008 2009

$ (mm)

Online searches 50% pa

Examples

Welcome! to your Natpal Account
Natpal

Select Segment

Electrical
Auto
Bakery
Banquet Room
Bridal Shop
Carer
Chiropractic
Computer Repair
Dentist
Dermatologist
Ecommerce Website
Electrician

Select Categories

- Accident
- Commercial
- 10/30 AMP Services
- Air Conditioning
- Appliances
- Electrical Installations
- Roof Cleaning
- Rewiring
- Residential Lighting
- Commercial & Industrial Cleaning
- Foam Upholstery

Save
Select Context

Select Budget
Keyword Creation

<table>
<thead>
<tr>
<th>Keywords Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>construction electrician philadelphia pa</td>
</tr>
<tr>
<td>electrician in philadelphia</td>
</tr>
<tr>
<td>construction electricians philadelphia</td>
</tr>
<tr>
<td>electrical contractor website in philadelphia</td>
</tr>
<tr>
<td>commercial electricians in philadelphia</td>
</tr>
<tr>
<td>construction electrician in philadelphia</td>
</tr>
<tr>
<td>certified electricians philadelphia pa</td>
</tr>
<tr>
<td>commercial electricians in philadelphia</td>
</tr>
<tr>
<td>electrical utility contractor philadelphia</td>
</tr>
<tr>
<td>commercial electricians philadelphia</td>
</tr>
<tr>
<td>electric help for systems philadelphia pa</td>
</tr>
<tr>
<td>construction electricians philadelphia pa</td>
</tr>
<tr>
<td>certified electricians philadelphia pennsylvania</td>
</tr>
<tr>
<td>certified electrician philadelphia</td>
</tr>
<tr>
<td>electrical contractor website philadelphia</td>
</tr>
<tr>
<td>electric help in philadelphia</td>
</tr>
</tbody>
</table>

Automatically Configures

Internet and Local

- Results from Even-Dar et al. (2007): Overall social welfare (advertiser and auctioneer surplus) increases when moving from standard to context-based auctions
  - However, one party can benefit at the expense of the other
- Broader Implications of Context-based Auctions
  - YP market is fundamentally changing
  - Many YPs have launched online YPs (superpages.com)
    - But search driving most of the traffic
Empirical Issues in Sponsored Search

Synthetic Bid Generation

- Ganchev et al. (2007) study bids placed in Yahoo’s (old) open auction
Traditional Distributions did not do well in generating realistic synthetic bids
- Observe significant jamming (bid differences of 1c)

Clicks Estimation
- Ali and Scarr (2007) estimate clicks using data from Yahoo
  - Use Negative Binomial, Zipf, Zipf-Mandelbrot, Zero-Altered, Zero-Inflated mixture models
  - Finding: Zipf-Mandelbrot (ZM) and Zero-Adjusted ZM are effective in modeling clicks distribution
Estimating CTR for new Ads

- Search engine needs to estimate CTR of new ad in order to place it
- Hard to estimate in the absence of historical data
- Richardson et al. (2007) run a logistic regression to estimate probability of click
  - Factors
    - CTR for other ads (for same keyword)
    - CTR for related terms (subset of superset of terms)
    - Ad Quality (appropriate capitalization, use of punctuation, etc)
  - Each factor helped improve fit by 20% or more

Conclusions

- Several Challenges and Open Questions
  - Mechanism Design
    - What are the best rules for ranking, payment
    - Open versus opaque auction
  - Bid Optimization
    - Combinatorial Problem
    - Stochastic
  - Keyphrase Generation
    - Machine Learning
  - Empirical Studies
    - Behavior Modeling: Estimating clicks, conversions
    - Estimating clicks and other functional relationships
Conclusions

- Search Engine Marketing is an interdisciplinary field
  - Economics (Mechanism Design)
  - Computer Science (Algorithms, Machine Learning)
  - Statistics (Resolving uncertainty)
  - Marketing (Modeling Consumer Behavior)

References

References