Introduction to Computational Advertising

ISM293
University of California, Santa Cruz
Spring 2009
Instructors: Ram Akella, Andrei Broder, and Vanja Josifovski
Questions about last lecture?

• We welcome questions & suggestions about all aspects of the course!
• E-mail to lsm293-group@soe.ucsc.edu
Computational advertising in the popular press …


ON a recent Thursday, Darren Herman, the president of Varick Media Management, was sequestered in his SoHo office. He wasn’t scrutinizing a television ad or images from a photo shoot. He was combing through graphs and Excel spreadsheets.

Mr. Herman had run 27 ads on the Web for his client Vespa, the scooter

http://www.nytimes.com/2009/05/31/business/media/31ad.html
Web analytics: Comparisons of display ad efficiency

Test Drive

Three online ads for Vespa performed very differently — as measured by clicks and other factors — over a short testing period. One ad was pulled, while another won more slots.

Test group: people in the market for station wagons.

Source: Varick Media Management

THE NEW YORK TIMES
Targeting
Disclaimers

- This talk presents the opinions of the authors. It does not necessarily reflect the views of Yahoo! Inc or any other entity.
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Yahoo! Or any other company.
Lecture plan for today

- Part I – Targeting
  - Demographic targeting
  - Behavioral Targeting
  - Re-targeting
- Part II – Panel, feedback, etc
Demographic Targeting
Using demographics in advertising

- Important indicator of people’s interest and potential of a conversion
  - Imagine you want to sell a $50K sports car. Who do you target?
- Used widely in traditional advertising:
  - TV, magazines, etc. maintain very detailed statistics of their audience
- Common dimensions:
  - Age
  - Gender
  - Income bracket
  - Location
  - Interests (“Golf enthusiast”)
  - ....
- Each dimension multiple values
Use of demographic information

1. **Main use: Targeting** of display advertising
   - Advertisers buy slices of the traffic based on demographic information
     - e.g. males, 25-35, California
     - Demographics
   - Impressions without associated demographic information cannot be sold in these cases!

2. **Secondary use: Improved textual ad selection**
   - Improve CTR/CPA.
   - In addition to other features (context, query, history)
   - Backfill in places with no immediate context, e.g. low content web pages in content match
   - Some targeting dimension specifiable also in textual advertising, e.g. geo.
Obtaining Demographic Information

- User supplied demographic information
  - Most reliable – if filled correctly
    - In some cases 15-20% of users born on 1\textsuperscript{st} of January
  - Most users see very little incentive to fill the form
  - Privacy concerns
  - But credit card data, shipping address, etc are almost 100% reliable.

- Inferred demographic information
  - Guess the demographics based on user browsing/querying behavior
    - 74% women/ 58% of men seek health or medical info online
    - 34% women/ 25% men seek religious info online
  - Wider reach – virtually every user
Inferring demographics

- How to infer the demographics from past behavior?
- **Classification**: e.g. regression model on the top of features extracted from history

- **Bipartite graph approach**:
  - Analyze the bipartite graph of users and their web pages/searches
  - Seed the graph with some demographic information
  - Infer demographics of users without the info

- **Combined approach**
  - ‘Demographic Prediction Based on User’s Browsing Behavior’. Hu et al, WWW 2007
Bipartite graph: Users and Web page visits
Bootstrapping: Assign attributes to the web pages

- Each web page assigned tendency – probability distribution over the space of possible demographics attributes (gender & age)
- Training set construction:
  - Textual features extracted from the pages
  - Training labels based on the available user information:
    \[
    \Pr(c \mid w_j) = \frac{\sum_{i=1}^{I} r_{ij} u_i(c)}{\sum_{i=1}^{I} r_{ij}}
    \]
  - Feature selection based on
    - Distribution grade on pages (Only discriminative pages kept)
    - Information gain for words on pages.
  - One binary classification SVM for each demographic dimension value
  - Normalize the outputs for each demographic dimension to translate to probabilities
- Smooth the page probability based on similar page by content
  - Reduced dimensional space (LSI – based on SVD)
  - The original feature space
From pages to users

- Simple Bayesian mapping of the page labels to the unlabeled users:
  - \( P(C|u_i) = \prod p(C|p_j) \), where \( p_j \) are pages connected to user \( u_i \)

- Smoothing can be applied on the user side as well, using the procedure

- Result:

![Bar chart showing performance for user gender/age prediction of CF, LSI, User side classifier, and our approach.](image)
F-measure in IR

- F = Harmonic mean of recall and precision
  \[ 2 \times \text{Pre} \times \text{Rec} / (\text{Rec} + \text{Pre}) \]
- Micro-average F
  = weights equally all samples (weights more heavily the popular categories)
- Macro-average F:
  = weights equally all categories (average of per-category averages)
Conclusion: Demographic Targeting

- Demographic targeting is widely used in traditional media
- A must for any display or textual ad network
- How to get the user demographic information?
- Infer demographics from user activity
- Combination of bipartite graph traversal and features of the users and page:
  - Seed with the entered user information
  - Similar to the state-of-the-art recommender systems
  - Smoothing in the features space and the reduced dimension space
- Issue: how much can we trust the entered user information?
Geo targeting

- Goal: determine user location
  - Home
  - Current
  - Often wrong 😞
- Inputs
  - Registration data
  - IP
  - Browser default language
  - Search language
  - Etc …
Behavioral Targeting
Ads are more influential than you might think! (2/3 of online users reacted to some ad)

<table>
<thead>
<tr>
<th>Action</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entered a sweepstakes</td>
<td>33%</td>
</tr>
<tr>
<td>Visited Web site related to ad</td>
<td>32%</td>
</tr>
<tr>
<td>Used search to find out more about product/service</td>
<td>27%</td>
</tr>
<tr>
<td>Purchased product/service online</td>
<td>21%</td>
</tr>
<tr>
<td>Registered with site for coupon/discount</td>
<td>17%</td>
</tr>
<tr>
<td>Played a game online</td>
<td>17%</td>
</tr>
<tr>
<td>Purchased in a physical store</td>
<td>14%</td>
</tr>
<tr>
<td>Told a friend about ad/product</td>
<td>12%</td>
</tr>
<tr>
<td>Forwarded ad to a friend</td>
<td>8%</td>
</tr>
<tr>
<td>Called toll-free to order</td>
<td>7%</td>
</tr>
<tr>
<td>None of the above</td>
<td>34%</td>
</tr>
</tbody>
</table>

66% of Online Users

Source: AOL/AudienceScience/JupiterResearch Consumer Survey (04/07)
What is BT?

- A technique used by publishers and advertisers to increase campaign effectiveness based on a given user’s historical behavior:
  - Previous searches/search sessions
  - Previous browsing activity
  - Previous ad-cliks
  - Previous conversions
  - Declared demographics data
  - Etc.

- Utility – everyone wins! (at least in theory 😊)
  - Advertisers: get a more appropriate/receptive audience, increased conversion rate, better ROI
  - Publishers: can ask for a premium
  - Users: see more interesting ads
Flavors

• **On-site BT**
  - Used mostly by e-commerce sites
  - Sometimes very elementary:
    - “Last three gizmos you looked at”
  - More sophisticated systems use recommender systems techniques
  - We are not talking about this …

• What we are talking about is **Network BT**
  - Information about the user is captured in one context and used for advertising in other contexts/situations, possibly on different web sites.
How popular is BT?

Do you currently use or plan to use display advertising?
- yes: 59.7%
- no: 26.2%
- unsure: 14.1%

Do you currently use or plan to use behavioral targeting?
- yes: 65.0%
- no: 23.2%
- unsure: 11.8%

Do you believe behavioral targeting is effective?
- yes: 74.9%
- no: 6.1%
- unsure: 19.0%

Source: Datran Media survey of 3,000 execs from Fortune 1000 companies December 2008
Papers

- Chen & al., Large-Scale Behavioral Targeting, KDD 2009
Conclusions of Yan & al.:

1. Users that click on the same ads have highly similar past behavior profile
   ➔ Profile can be used to predict CTR
2. User can be clustered based on profile (unsupervised learning). On a 160 cluster partition, one can observe a 600% prediction improvement. (Sort of. Will explain in a moment! 😊)
3. Short term behavior profiles are more predictive than long term behavior profiles.
Data collection

- ~ 6M users, ~ 17K ads
- Long term = 7 days, Short term = 1 day
- Browse behavior captured in $g \times u$ matrix
  - $g$ users
  - $u$ URLs
  - TFIDF approach to visited URLs (lognormal)
- Search behavior captured in $g \times h$ matrix
  - $g$ users
  - $h$ query terms
  - TFIDF approach to visited query terms
Similarity

- “Within ads” similarity:

\[ S_w(a_i) = \frac{2}{l_i(l_i - 1)} \sum \sum Sim(u_{ij}, u_{it}) \]

- “Between ads” similarity

\[ S_b(a_i, a_s) = \frac{1}{l_i l_s} \sum \sum Sim(u_{ij}, u_{st}) \]

- Dissimilarity Ratio

\[ R(a_i, a_s) = \frac{S_w(a_i) + S_w(a_s)}{2S_b(a_i, a_s)} \]
Observed within-ads & between-ads similarity & ratios

- Define
  1. Average within-ad similarity (over all ads)
  2. Average between-ad similarity (over all ad pairs)
  3. Average between-ad \( R \) (over all ad pairs)

- Yan & al. observed numbers:

  Within- and between- ads user similarity.

<table>
<thead>
<tr>
<th></th>
<th>( S_w )</th>
<th>( S_b )</th>
<th>( R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>0.1417</td>
<td>0.0252</td>
<td>28.9217</td>
</tr>
<tr>
<td>LQ</td>
<td>0.2239</td>
<td>0.0196</td>
<td>44.2908</td>
</tr>
<tr>
<td>SP</td>
<td>0.1532</td>
<td>0.0281</td>
<td>24.5086</td>
</tr>
<tr>
<td>SQ</td>
<td>0.2594</td>
<td>0.0161</td>
<td>91.1890</td>
</tr>
</tbody>
</table>
User segmentation, precision, and recall

- Assume that we partition the users into a set segments \( \{g\} \)
- For a given ad and a given segment define the CTR improvement

\[
\Delta(a_i) = \frac{CTR(a_i|g_*(a_i)) - CTR(a_i)}{CTR(a_i)}
\]

where \( g_* \) is the “best” segment (highest CTR) for \( a_i \)
- Then we can define an average improvement over all ads.
CTR “improvements”

User clustering by k-means
Entropy reduction

\[ Enp(a_i) = - \sum_{k=1}^{K} P(g_k|a_i) \log P(g_k|a_i) \]

\[ = - \sum_{k=1}^{K} \frac{1}{K} \log \frac{1}{K} = \log K \]
My conclusions

- The behaviour profile (in particular search) carries some signal
- Short term is more indicative than long term
- Cold start remains a problem – what do we do about new ads?
- We should look at high powered recommendation system techniques
Search Re-targeting
Re-targeting idea

- Use immediate search or browse to target/create ads
  - Examples:
    1. User that has searched for “Prius” sees ads for Prius or Toyota dealer for the next few days on non-search context, e.g. when browsing a complete different site
    2. User that has browsed a fashion site sees ads for shoes when browsing a complete different site or using e-mail, etc
    3. User that has browsed BuyAGizmo.com but did not convert, sees “get back” ads for BuyAGizmo.com on many other sites, maybe + coupon, special discount

- Mostly search re-targeting due to higher intentionality
- Special case of BT (more recent, more specific)
- Companies: Advertising.com, FetchBack, Real Media, Dapper, Microsoft DrivePM, Audience Science, BlueLithium (Yahoo!),...
Basic search retargeting scheme

- User searches for shoes on the XYZ engine

Site ABC sends ad request + XYZ cookie to XYZ

XYZ creates shoes ad based on XYZ cookie that remembers “shoes”
Basic browse retargeting scheme

- User Joe views skiing site ABC that contains some XYZ produced ad or just “beacon”
- Joe’s XYZ cookie captures visit to ABC
- Now on site DEF Joe sees ski ads

Sends ad request + XYZ cookie to XYZ

- XYZ creates skiing ad based on XYZ cookie that remembers “skiing”
- Alternative: XYZ puts ad for ABC
Example: dapper.net

Behavioral Remessaging
AOL explanation

Did you know that many ads you see on the web are based on other websites you have visited? An ad company sends a cookie to Mr. Penguin’s computer, recording his visit.


The ad company reads the cookie to display a relevant ad.

Mr. Penguin visits AnchovyGourmet.com

For more information from AOL about online advertising and your privacy choices: Click Here.
Data exchanges

- **Publisher (seller)**
  - Collects behavioral data on content or e-commerce site
  - And/or collects straight demographic data

- **Sell**
  - To data exchange → Resells to advertisers
  - Directly to advertiser

- **Advertiser (buyer)**
  - Buy impressions directly from another publisher via an ad network or ad exchange.
  - Uses the bought data for targeting and/or buying

- All done with cookies and beacons
- See discussion at http://adage.com/adnetworkexchangeguide09/article?article_id=136003
- **Key Quote:**
  - Aggregators and Exchanges Aim to Create 'Liquid Market' Based on Users' Activities, Not Their Locations -- but Can They Get Past Privacy Concerns?
- Companies: BlueKai, eXelate, Datran, Lotame, SocialMedia, Media6Degrees, …
Privacy concerns

- Users do not understand the cookie mechanisms
- Difficult to turn off – many sites stop being functional without cookies
- If you accept cookies from XYZ, XYZ can become aware of your visits to any site where XYZ has a visible or invisible presence on the page.
- Many proposals / regulations / “trust-me” solutions
  - E.g. Phorm (http://www.phorm.com/) promises to collect only category data and keep cookies anonymous (not linked to ip, name, etc)
  - Most companies have data retention policies
  - Most companies allow user control over stored data
  - Opt-out BT
  - Etc
Key points

- Targeting is emerging as a key component of online advertising
- Demographic targeting mimics classic advertising, but it can be both declared and inferred
- Behavioral Targeting based on recent search and query behaviour appears effective, quantitative aspects still TBD
- Re-targeting is becoming very popular, with new players mushrooming
- Privacy and regulatory concerns
Thank you!

broder@yahoo-inc.com
vanjaj@yahoo-inc.com

http://research.yahoo.com
This talk is Copyright 2009. Authors retain all rights, including copyrights and distribution rights. No publication or further distribution in full or in part permitted without explicit written permission.