Targeting Converters in Display Advertising

Deepak Agarwal, Yahoo! Research
Joint with Sandeep Pandey, Y! Research

University of Texas, Austin
15th Nov, 2011
Agenda

• Background on Advertising
• Display Advertising
  – Futures Market
  – Spot Market --- Ad-exchange, Real-time bidder (RTB)
• User Targeting Problem in Display Advertising
  – Current practice, state-of-the-art

• Our approach
  – Using historic campaign data to target new campaigns

• Modeling Details
  – Factor Model

• Experiments
The two basic forms of advertising

1. **Brand advertising**
   - creates a distinct favorable image

2. **Direct-marketing**
   - Advertising that strives to solicit a "direct response":
     - buy, subscribe, vote, donate, etc, **now or soon**
Brand advertising ...

60 DAYS OF DAYLIGHT FOR APARTMENT 6F.

SMITHWICK’S IS BRINGING THE LUCK OF THE IRISH TO NYC.

Good luck comes to those who drink responsibly.

Ireland's Oldest Ale.
Sometimes both Brand and Performance
Web Advertising

There are lots of ads on the web ... 100s of billions of advertising dollars spent online per year (e-marketer)
Web Advertising: Comes in different flavors

• Sponsored (“Paid”) Search
  – Small text links in response to query to a search engine

• Display Advertising
  – Graphical, banner, rich media; appears in several contexts like when visiting a webpage, checking e-mails, facebook, ….
    • Brand Awareness campaign
    • Performance campaign
      – More like direct marketing in offline world, target users for some favorable response in the near-term (buy, vote, etc)
Paid Search: Advertise Text Links

Search for "hotels near disneyland" on Google, showing ads for hotels near Disneyland.

- Hotels Near Disneyland | Marriott.com
- Hotels Near Disneyland | HolidayInn.com
- Hotels Near Disneyland | Booking.com
- Good Neighbor Hotels Near Disneyland® Resort | Expedia.com
- Disneyland.disney.go.com/hotels/good-neighbor
- Anaheim Hotel
- Affordable Anaheim Hotels Near Disneyland
- 145 Anaheim Great Hotels

Deepak Agarwal @UTAustin’11
Display Advertising: Examples

- Stocks surge as Italy, Greece allay debt fears AP
- SEC disciplines 8 employees over Madoff failure AP
- CME offers $300M to help unfreeze MF Global funds AP
- Universal, Sony/ATV to buy EMI for $4.1 billion AP
- NPD: October video game sales rise 1 percent AP
- Fed to conduct 4th round of bank stress tests AP
- Electric car battery catches fire after crash test AP
- Italy moves toward economic and political change AP

Currencies

<table>
<thead>
<tr>
<th>Name</th>
<th>Price</th>
<th>Change</th>
<th>% Chg</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/USD</td>
<td>1.3750</td>
<td>+0.0143</td>
<td>+1.65%</td>
</tr>
<tr>
<td>USD/JPY</td>
<td>77.3835</td>
<td>-0.2680</td>
<td>-0.35%</td>
</tr>
<tr>
<td>AUD/USD</td>
<td>0.9842</td>
<td>-0.0013</td>
<td>-0.13%</td>
</tr>
<tr>
<td>CAD/USD</td>
<td>0.9740</td>
<td>+0.0017</td>
<td>+0.18%</td>
</tr>
<tr>
<td>NZD/USD</td>
<td>0.7700</td>
<td>-0.0033</td>
<td>-0.43%</td>
</tr>
<tr>
<td>HKD/USD</td>
<td>15.2024</td>
<td>-0.0238</td>
<td>-0.15%</td>
</tr>
<tr>
<td>RUB/USD</td>
<td>33.7135</td>
<td>-0.0033</td>
<td>-0.09%</td>
</tr>
<tr>
<td>THB/USD</td>
<td>23.3634</td>
<td>+0.0877</td>
<td>+0.37%</td>
</tr>
<tr>
<td>INR/USD</td>
<td>69.0645</td>
<td>+0.1540</td>
<td>+0.22%</td>
</tr>
<tr>
<td>IDR/USD</td>
<td>16253.00</td>
<td>-0.0008</td>
<td>-0.00%</td>
</tr>
<tr>
<td>MYR/USD</td>
<td>3.9460</td>
<td>-0.0004</td>
<td>-0.01%</td>
</tr>
<tr>
<td>TWD/USD</td>
<td>30.9000</td>
<td>-0.0174</td>
<td>-0.06%</td>
</tr>
<tr>
<td>BGN/USD</td>
<td>1.9216</td>
<td>-0.0001</td>
<td>-0.01%</td>
</tr>
<tr>
<td>CZK/USD</td>
<td>27.4878</td>
<td>+0.0001</td>
<td>+0.00%</td>
</tr>
<tr>
<td>HRK/USD</td>
<td>0.7301</td>
<td>-0.0001</td>
<td>-0.01%</td>
</tr>
<tr>
<td>HUF/USD</td>
<td>289.1333</td>
<td>-0.0012</td>
<td>-0.04%</td>
</tr>
<tr>
<td>PLN/USD</td>
<td>3.6317</td>
<td>+0.0005</td>
<td>+0.01%</td>
</tr>
<tr>
<td>SEK/USD</td>
<td>8.5021</td>
<td>+0.0034</td>
<td>+0.40%</td>
</tr>
<tr>
<td>NOK/USD</td>
<td>7.4131</td>
<td>+0.0034</td>
<td>+0.45%</td>
</tr>
<tr>
<td>DKK/USD</td>
<td>5.7496</td>
<td>+0.0031</td>
<td>+0.05%</td>
</tr>
<tr>
<td>CHF/USD</td>
<td>0.9703</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>114.2603</td>
<td>+0.0009</td>
<td>+0.08%</td>
</tr>
<tr>
<td>KRW/USD</td>
<td>1126.37</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>RON/USD</td>
<td>0.3833</td>
<td>+0.0003</td>
<td>+0.76%</td>
</tr>
<tr>
<td>RUB/USD</td>
<td>33.7135</td>
<td>-0.0033</td>
<td>-0.09%</td>
</tr>
<tr>
<td>HUF/USD</td>
<td>289.1333</td>
<td>-0.0012</td>
<td>-0.04%</td>
</tr>
<tr>
<td>PLN/USD</td>
<td>3.6317</td>
<td>+0.0005</td>
<td>+0.01%</td>
</tr>
<tr>
<td>SEK/USD</td>
<td>8.5021</td>
<td>+0.0034</td>
<td>+0.40%</td>
</tr>
<tr>
<td>NOK/USD</td>
<td>7.4131</td>
<td>+0.0034</td>
<td>+0.45%</td>
</tr>
<tr>
<td>DKK/USD</td>
<td>5.7496</td>
<td>+0.0031</td>
<td>+0.05%</td>
</tr>
<tr>
<td>CHF/USD</td>
<td>0.9703</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>114.2603</td>
<td>+0.0009</td>
<td>+0.08%</td>
</tr>
<tr>
<td>KRW/USD</td>
<td>1126.37</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>RON/USD</td>
<td>0.3833</td>
<td>+0.0003</td>
<td>+0.76%</td>
</tr>
<tr>
<td>RUB/USD</td>
<td>33.7135</td>
<td>-0.0033</td>
<td>-0.09%</td>
</tr>
<tr>
<td>HUF/USD</td>
<td>289.1333</td>
<td>-0.0012</td>
<td>-0.04%</td>
</tr>
<tr>
<td>PLN/USD</td>
<td>3.6317</td>
<td>+0.0005</td>
<td>+0.01%</td>
</tr>
<tr>
<td>SEK/USD</td>
<td>8.5021</td>
<td>+0.0034</td>
<td>+0.40%</td>
</tr>
<tr>
<td>NOK/USD</td>
<td>7.4131</td>
<td>+0.0034</td>
<td>+0.45%</td>
</tr>
<tr>
<td>DKK/USD</td>
<td>5.7496</td>
<td>+0.0031</td>
<td>+0.05%</td>
</tr>
<tr>
<td>CHF/USD</td>
<td>0.9703</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>114.2603</td>
<td>+0.0009</td>
<td>+0.08%</td>
</tr>
<tr>
<td>KRW/USD</td>
<td>1126.37</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>RON/USD</td>
<td>0.3833</td>
<td>+0.0003</td>
<td>+0.76%</td>
</tr>
<tr>
<td>RUB/USD</td>
<td>33.7135</td>
<td>-0.0033</td>
<td>-0.09%</td>
</tr>
<tr>
<td>HUF/USD</td>
<td>289.1333</td>
<td>-0.0012</td>
<td>-0.04%</td>
</tr>
<tr>
<td>PLN/USD</td>
<td>3.6317</td>
<td>+0.0005</td>
<td>+0.01%</td>
</tr>
<tr>
<td>SEK/USD</td>
<td>8.5021</td>
<td>+0.0034</td>
<td>+0.40%</td>
</tr>
<tr>
<td>NOK/USD</td>
<td>7.4131</td>
<td>+0.0034</td>
<td>+0.45%</td>
</tr>
<tr>
<td>DKK/USD</td>
<td>5.7496</td>
<td>+0.0031</td>
<td>+0.05%</td>
</tr>
<tr>
<td>CHF/USD</td>
<td>0.9703</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
<tr>
<td>JPY/USD</td>
<td>114.2603</td>
<td>+0.0009</td>
<td>+0.08%</td>
</tr>
<tr>
<td>KRW/USD</td>
<td>1126.37</td>
<td>-0.0003</td>
<td>-0.03%</td>
</tr>
</tbody>
</table>
Display Advertising: Examples

- An example of display advertising on a Yahoo! Mail page.

Deepak Agarwal @UTAustin’11
Display Advertising: Examples

哂

Liz Morgan

Hi Deepak - I am part of the Online Services Division with Microsoft. Our product portfolio includes Bing, Bing Mobile, MSN, Ads Platform, and the Global Foundation Services group.

http://www.microsoft-careers.com/content/bing/bing-home/

Would you be interested in discussing potential Researcher opportunities with this division? We have several openings that have an emphasis on AI, machine learning, and data mining.

Regards,
Liz Morgan

Attachment Unavailable

The attachment source was deleted or the privacy settings on this attachment do not allow you to view it.

Write a reply...

Pickle Addicts

PIKLES! You know you want them! Visit our FB page for all things pickles! Like us for special FB only pdole offers!

Like · 1,294

Facebook © 2011 · English (US)

About · Advertising · Create a Page · Developers · Careers · Privacy · Terms · Help
Paid Search versus Display Advertising

**Paid Search**
- Context (Query) important to match
- Small text links
- Performance based
  - Clicks, conversions
- Advertisers can cherry-pick *instances*

**Display**
- Reaching desired audience main goal
- Graphical, banner, Rich media
  - Text, logos, videos,..
- Hybrid
  - Brand, performance
- Bulk buy
  - But things evolving
    - Ad exchanges, Real-time bidder (RTB)
Display Advertising Models

• Futures Market (Guaranteed Delivery)
  – Brand Awareness (e.g. Coke, MacDonald’s,..)

• Spot Market (Non-guaranteed)
  – Marketers create campaigns targeted to user segments
    • Ad-exchanges have made this process efficient
      – Connects buyers and sellers in a stock-market style market
Guaranteed Delivery (Futures Market)

- Revenue Model: Cost per eyeball (CPM)

- Traditional Advertising Model:
  Ads are bought in bulk targeted to users based on demographics and other behavioral features
  - GM ads in Y! autos shown to “males above 55”
  - Mortgage ad shown to “everybody on Y! Front page”

Book a slot well in advance
- “2M impressions in Jan next year”
- Future impressions must be guaranteed by the ad network
- Prices are significantly higher than in Spot market
  - Publishers ensure higher quality inventory
Spot Market

- Different advertiser objectives
  - Brand awareness to target future sales: CPM
  - Performance based (auto, washing machine, pizza, donation,...)
  - Pay by conversion (favorable action on landing page) or/and click

Advertisers

Car Insurance

Sports

Accessories

Online

Education

submit ads to the network

display ads for the network

Intermediaries

www.cars.com

www.sportsauthority.com

www.elearners.com

AD-EXCHANGE

Publishers

Deepak Agarwal @UTAustin’11
Advertising Eco-System

• Ad exchange (like a stock-market)
  – Winner decided through an auction process
    • Several intermediaries (ad-networks, pub-networks,…)
  – Still bulk buy, advertisers specify targets and a bid

• New technologies
  – Real-time bidder: change bid dynamically
    • Based on user cookie and other information passed by publishers to the broker
      – New intermediaries: sell user data (BlueKai,…)
  – Demand side platforms: single unified platform to buy inventories on multiple ad-exchanges
"Half the money I spend on advertising is wasted; the trouble is, I don't know which half." - John Wanamaker

PROBLEM: BETTER USER TARGETING
Our problem: User Targeting for Performance

- Given a campaign that is focused on conversions, how do we target the best users to maximize advertiser ROI and publisher revenue?

- Why conversions? Why not clicks?
  - Conversions better and direct measure of user intent to buy
  - Clicks subject to problems: Fraud, bounce
  - Advertiser measure performance via conversions

- One important aspect of conversions
  - Definition varies across campaigns
    - Buying, filling-up a survey, subscription, and so on.
Different kinds of user information

user

Demographics (Age, Gender)  behavioral  social  purchase  ……
Behavioral User profile: Main focus in this work

User feature vector

Queries
- lady gaga
- ipad
- harry potter
- angry birds

Page Views
- facebook.com
- music.yahoo.com
- blockbuster.com
- youtube.com

Ad views/clicks
- arcade games
- download itunes
- overstock ipads
- buy ipad now

Text extraction
Campaign information

Campaign metadata

Campaign

Creative 1

Creative 2

Creative 3

landing page

landing page

other campaigns
Typical targeting

- Based on campaign information, look at different user dimensions and create a few targeting segments
  - E.g. campaign to sell Halloween costumes in Austin
  - Targeting attributes:
    - Users 25-50 living around Austin who searched for Halloween
    - Subjective, depends on the marketer

- Experimentation
  - Run the campaign for a while, collect conversion data and build a model to target users in a better way
    - Conversion rates low, experiment is expensive
    - Advertisers often lose patience and shut-off poor performing campaigns
Our Approach

• We take a fundamentally different approach
• Lots of campaigns run everyday in ad-exchanges, can we leverage this data and perform better user targeting for future campaigns?
  – Such data can also be constructed by other intermediaries in the advertising eco-system
• Intuitively, this seems plausible
  – Create clusters of campaigns, figure out where a new campaign fits
• Statistically, this is a challenging task
Our Approach: Use past campaign performance

TRAINING

Users
- profile
  - pages visited
  - search queries
  - ad views

Campaigns
- metadata
  - ad creatives
  - landing pages

MODELS:
- data pooling across campaigns (REG, FACTOR)
- underlying structure (coefficient matrix, factors)

INFERENCES

- initial model (user features and weights)
- refined model
- converted/non-converted users
- new campaign metadata

Deepak Agarwal @UTAustin’11
MODELING
Per campaign model

- \( y_{kj} \) \( k^{th} \) targeted user on campaign j (convert/non-convert)
- \( X_{kj} \) user feature vector (high dimensional)
- Model: Logistic regression

\[
y_{kj} \sim \text{Bernoulli}(p_{kj}); k \in \mathcal{U}_j
\]
\[
\log \frac{p_{kj}}{1-p_{kj}} = X'_{k,j} \beta_j
\]

• Difficulties
  – Sparseness & rareness of conversions
    • Hundreds of thousands of user features, few hundred to few thousand conversions per campaign
User Feature distributions: all features are binary (~100k)

- Feature occurrence across campaigns
- Feature sparsity within campaign
Logistic regression ill-conditioned

- Need: Feature cones of converters and non-converters to overlap (Silvapulle, 1981)
- Difficult to achieve
  - Hundreds of thousands of features, most of them sparse
  - Few hundred to few thousand conversions

- Approach: Multi-level Hierarchical models
  - Assume the coefficients are drawn from some prior distribution

\[ \beta_{ij} \sim N(\beta_{ij0}, \sigma^2) \]
Modeling Prior Mean

• **ZEROMEAN**: Usual L2 regularization

\[ \beta_{ij0} = 0 \]

  – State-of-the-art in Display Targeting
  – Limitation: Does not work with little or no data (Cold Start)

• **REG**

\[ \beta_{ij0} = g_i' z_j = \sum_k g_{ik}z_{jk} \]

\[ G = (g_1, g_2, \ldots) \]

  – Too many parameters: e.g. \(100k \times 30 = 3M\)
  • Use L2 regularization on \(G\)
New Approach: FACTOR

- Factorize the incomplete feature-id x campaign matrix
- Idea borrowed from collaborative filtering
  - Factorize User x Item matrix
  - Here factorization one level up in the multi-hierarchy model

\[
\beta_{ij0} = u_i' v_j \\
u_i \sim MVN(0, a_u) \\
v_j \sim MVN(Dz_j, a_v)
\]

- \( r = \# \text{ factors (small, e.g. 4-5)} \)
- \( D : (r + \#\text{campaign features}) \) (small compared to REG)
Advantages of FACTOR

- Factorizing coefficient matrix reduces dimension
- Alternate parameterization of FACTOR

\[
\beta_{ij0} = u_i' (Dz_j + \eta_{jrx1}) + \epsilon_{ij} \\
= u_i' Dz_j + u_i' \eta_{jrx1} + \epsilon_{ij}
\]

Sharing parameters

- Recall REG

\[
\beta_{ij0} = g_i z_j + \epsilon_{ij}
\]
Other related approaches

• Seemingly Unrelated Regressions (SUR) [Zellner, 1962]
  – We deal with logistic + model covariance in residuals through campaign features and low-rank matrix structure

• Multi-tasking
  – Assumes complete feature-id x campaign matrix (not the case)
  – Does not incorporate campaign feature information
  – Usual approach assumes
    \[ \beta_{ij0} = \mathbf{u}_i^* \eta_j \text{ (Low Rank)} \]
    – low-rank + sparse (we assume low-rank + white-noise)
      • Provides robust estimates after adjusting for heterogeneity
      – Relaxation important to adapt estimates with online data for new campaign, robustness is not the only issue here
Model Fitting
EM for our models (REG and FACTOR)

\[ Y : \text{Data} \]

\[ \Delta : \text{Latent variables} \]

\[ \Theta : \text{hyper-parameters} \]

Model: \[ p(Y|\Delta, \Theta)p(\Delta|\Theta) \]

Output needed: Mode: \[ \max_{\Theta} p(\Theta|Y) \]

\[ p(\Delta|Y) \approx p(\Delta|Y, \hat{\Theta}) \]
Computing the mode

\[
\log(p(\Theta|Y)) = \log(p(\Theta, \Delta|Y)) - \log(p(\Delta|\Theta, Y))
\]

\[
\log(p(\Theta|Y)) = E_{old}(\log(p(\Theta, \Delta|Y))) - E_{old}(\log(p(\Delta|\Theta, Y)))
\]

\[E_{old} : \text{Expectation w.r.t. } p(\Delta|\Theta_{old}, Y)\]

Second term: Maximized at \(\Theta_{old}\)
Find new value of \(\Theta\) that increases first term
The EM algorithm

Initialize $\Theta$

Iterate

E-step : $E_{old}(\log(p(\Theta, \Delta | Y)))$

M-step : $\text{argmax}_\Theta E_{old}(\log(p(\Theta, \Delta | Y)))$

Improving the log-likelihood in M-step is enough, no need to perform complete optimization (Generalized EM)
EM algorithm to fit REG

\[\Delta = \{\beta_{ij}\}_{i,j} \quad \Theta = (G, \sigma^2)\]

Posterior independence across campaigns in E-step

\[\left[\beta_1, \ldots, \beta_j, \ldots \mid Y_1, \ldots, Y_j, \ldots, \Theta\right] = \prod_j \left[\beta_j \mid Y_j, \Theta\right]\]

Posterior not in closed form

\[2 \ln[\beta_j \mid Y_j, \Theta] = 2 \sum_k (y_{kj} X'_{kj} \beta_j - l \exp(X'_{kj} \beta_j)) - \sum_i \left(\beta_{ij} - g_i z_j\right)^2 / \sigma^2 - C \log \sigma^2\]

We can sample through Gibbs sampling, each conditional log-concave
Adaptive Rejection Sampling Sampling (ARS) (Gilks and Wild, ’92)
Adaptive Rejection Sampling (Gilks and Wild)

- Create upper (and lower bounds) to the density
- Perform rejection sampling, refine the envelope with each rejection (adaptive)
EM algorithm for REG

- **E-step**: Sample using MCMC, parallelize on Hadoop
  - Mapper: split data by campaign
  - Reducer: run MCMC for each campaign

\[-2E \log p(\Delta, Y \mid \Theta) = \sum_{(i,j)} ((\hat{\beta}_{ij} - g'_i z_j)^2 + \tau_{ij}^2) / \sigma^2 + C \log \sigma^2\]

- **M-step**:
  - Estimate the $g$’s per feature-id through regression
  - Variance estimate is closed-form

\[C \hat{\sigma}^2 = \sum_{(i,j)} (\hat{\beta}_{ij} - \hat{g}'_i z_j)^2 + \tau_{ij}^2\]
EM for FACTOR

• Parameters $\Delta = (\{\beta_{ij}\}_{(i,j)}, \{u_i\}_i, \{v_j\}_j); \Theta = (D, \sigma^2, a_u, a_v)$

• Gibbs sampling for approximating posterior

  $[\{\beta_{ij}\}_{(i,j)} | \{u_i\}, \{v_j\}, REST]$: ARS (parallelize)

  $[u_I | \{u_i\}_{-I}, REST] = [u_I | REST]$ (parallel)

  $[v_J | \{v_j\}_{-J}, REST] = [v_J | REST]$ (parallel)

• Not amenable to Hadoop
  – One map-reduce job per MCMC iteration, infeasible
Approximate EM for FACTOR

- **E-step:** Assume the factors to be fixed and sample the logistic coefficients using per campaign Gibbs sampler.

- **M-step:** Fit the following model to obtain estimates of $(\{u_i\}, \{v_j\}, D)$

  \[
  \hat{\beta}_{ij} \sim N(u_i' v_j, \sigma^2) \\
  u_i \sim MVN(0, a_u I); v_j \sim MVN(Dz_j, a_v I)
  \]

- \(\sigma^2\) estimated as in REG with change in prior mean to \(u_i v_j\)
Experiments
Data example

• 115 conversion based campaigns from Y! RightMedia
• 90 used for training, rest for testing
• TEST: simulate performance during different stages of a campaign’s lifetime

• Evaluation metric
  – AUC (Area under the ROC curve)
    • Pr( converter score > non-converter score)
ZEROMEAN performance

Area under the ROC curve

# positives examples in training (log scale)

weighted avg
unweighted avg
Comparing all methods

Weighted Average AUC

# positives examples in training (log scale)
Variation across campaigns

(a) ZEROMEAN

(b) REG

(c) FACTOR
Discussion

• Using previous campaign data via FACTOR helps provide better performance at the beginning of the campaign itself, encouraging for advertiser ROI

• FACTOR adapts as more conversions *trickle in* and improves model performance for a campaign

• Fitting algorithm scales in a map-reduce framework
To Conclude

- Targeting users is at the heart of Display Advertising that involves ingesting user information from disparate sources
  - e-marketer estimates this to be an 80B industry growing at 17% p.a
- Current techniques are subjective and involves marketers inspecting few dimensions based on domain knowledge
- We showed using past conversion data can be a useful source to build effective targeting strategy for new campaigns, such data available to multiple entities in the advertising eco-system
- We achieved this through a new multi-tasking approach called FACTOR that handles sparsity, heterogeneity, and scales in a map-reduce framework