SCALABLE, PRIVACY-FRIENDLY, TARGETED INTERNET ADS

Claudia Perlich
Chief Scientist

Thanks to INFORMS for the 2009 Design Science Award
OUTLINE

Industry

The high level M6D approach

M6D data science for finding the right audience
In a nutshell: our customers are brands who ask us to find a good audience and run an online ad campaign for them.

Main player in social targeting for (non-premium) display advertisement.

Founded in 2008.

Growth > 20 Million Revenue in 2010.
## Display Advertising Targeting Landscape

<table>
<thead>
<tr>
<th>Advertisers</th>
<th>Media Buying Platforms</th>
<th>&quot;DSPs&quot;</th>
<th>Ad Exchanges</th>
<th>Ad Networks</th>
<th>Horizontal</th>
<th>Social &amp; Tools</th>
<th>Sharing Data</th>
<th>Audience Optimization</th>
<th>Audience Tools</th>
<th>Publishers</th>
<th>Ad Servers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omnicom Group</td>
<td>Omnicom/Webedge Group, TradingDesk</td>
<td>invite media*</td>
<td>doubleclick by Google</td>
<td>Yahoo!</td>
<td>Google Aol.</td>
<td>Facebook, ShareThis, *clearspring, gigya</td>
<td>Yield Optimization</td>
<td>PubMatic, Rubicon</td>
<td>AdMeld, FatTail, YieldBot, Scout, Analytics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPP</td>
<td>m+</td>
<td>rightmedia</td>
<td>AECN</td>
<td>Aol.</td>
<td>Aol.</td>
<td>Adad, ShareThis, ShareThis, ShareThis, ShareThis, ShareThis, ShareThis</td>
<td>AdMeld, FatTail, YieldBot, Scout, Analytics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publicis</td>
<td>IPG</td>
<td>TRUMPR</td>
<td>ContextWeb, AdBuyer</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td>24/7, Undertone, Networks, BurstMedia, Traffic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKQA</td>
<td>LBi</td>
<td>TUMRI</td>
<td>DataOptim</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LBi</td>
<td>crossing</td>
<td>quantcast</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdOpting</td>
<td>Optim</td>
<td>quantcast</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DataOptim</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMPs and Data Aggregators</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targeted/Audience</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Suppliers</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytics</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verification/Attribution</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdOps/Infrastructure</td>
<td>Bluekai</td>
<td>Bluekai</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td>Bidr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CURRENT AD SPENDING SEEMS DISPROPORTIONATE...

SHARE OF TIME IN A TYPICAL WEEK THAT US ADULTS SPEND WITH SELECT MEDIA* VS. SHARE OF US ADVERTISING SPENDING BY MEDIA, 2007

Note: *consumer media time excludes time spent using a mobile phone, watching DVDs or playing video games; Source: Forrester Research, “Teleconference: The US Interactive Marketing Forecast 2007-2012,” January 4, 2008
ONLINE ADVERTISING SPENDING BREAKDOWN (2009)

% OF 2009 FULL YEAR REVENUES

- Digital Video: 4%
- Display/Banner Ads: 22%
- Sponsorship: 2%
- Classifieds: 10%
- Search: 47%
- E-mail: 1%
- Lead Generation: 6%
- Rich Media: 7%

Total - $22.7 Billion

Source: IAB Internet Advertising Revenue Report
Conducted by PricewaterhouseCoopers and Sponsored by the Interactive Advertising Bureau (IAB)
April 2010
SOCIAL TARGETING FOR ONLINE ADVERTISING
Every **BRAND** has a unique Social Signature, determined by the collection of sites where current customers cluster.

Every **CONSUMER** can be scored against the brand’s Social Signature, based on his/her visits to those sites.
STEP 1
IDENTIFY BRAND’S SOCIAL SIGNATURE

We place pixels on your site to cookie current brand customers.

Our algorithm identifies and weights sites where the customers visit.
• Sites with the highest visit density determine the brand’s Social Signature.
STEP 2
MATCH CONSUMERS WITH BRAND SIGNATURE

We score prospects based on how well they match your brand’s Social Signature.

Close match = better prospect…more likely to convert.
**STEP 3**

**PLACE THE AD WHEREVER THE BROWSER MAY BE**

We find the best prospects in the best environments at whatever scale your campaign requires, and deliver the ads.

1. Prospects who match a brand’s Social Signature visit sites in the ad exchanges.
2. We buy impressions and serve ads to users with the highest scores, via the major ad exchanges.
3. Users respond to the ads and visit brand site.

---

1. **Prospects who match a brand’s Social Signature visit sites in the ad exchanges.**
2. **We buy impressions and serve ads to users with the highest scores, via the major ad exchanges.**
3. **Users respond to the ads and visit brand site.**
STEP 4

REFINE YOUR SIGNATURE BUY

We track results via our full feedback loop:

- Refine your Social Signature.
- Refresh each user’s score to optimize performance in real-time.
MAIN POINTS FOR THIS MORNING

1. Machine learning can be used as the basis for **EFFECTIVE, PRIVACY-FRIENDLY** targeting for online advertising

2. Important to consider carefully the **TARGET** variable used for training

3. Many other exciting analytic challenges
   - Bid optimization
   - Across brand optimization
   - Server side cookie consistency
   - Proof ad effectiveness
   - Customer site optimization
BROWSER INTERACTIONS

1. Browsing with one of our data partners

2. Shopping at one of our campaign sites

3. General browsing with display ads

If we win an auction we serve an ad

Browser interaction with ad (click)
1. **DOUBLY-ANONYMIZED BIPARTITE GRAPH**

For each browser, we collect from our data partners hashed tokens of URL’s previously visited.

**BrowserID:**
1234

**ContentIDs:**
abkcc
kkllo
88iok
7uiol

From this data, we induce a bipartite browser-content graph.
2. ADDING LABELS TO THE BIPARTITE GRAPH

Additionally, our brand partners provide information on what cookies have previously taken some brand-related action.*

From this data, we include labels in the bipartite browser-content graph.

*note: this is done separately for every brand we work with
3. THE TARGETING GOAL

Which of the browsers that have not previously taken a brand action are likely to do so?

\[ P(\text{conv}|\text{imp}, \varphi_{bi})^* \]

\[ P(\text{conv}|\text{imp}, \varphi_{bi}) \]

In effect, this is a standard binary classification task.

*\(\varphi_{bi}\) is a set of features that capture unconditional brand affinity based on graph structures...next section
KEY MODELING CHALLENGES

DATA LIMITATIONS
No contextual information
No demographic or PII
Conversion rates ~ .01%

DIMENSIONALITY & DATA SIZE
> 100 MM Browsers
> 50 MM hashed URL’s
> 150 brands
data is very sparse

OPERATIONAL CONSTRAINTS
Training models in linear time
Scoring browsers in ms

How do you find a signal when the nature of the ecosystem and self regulation limits your data availability?

How do you operate efficiently on so much data for so many clients?
OUR ITERATIVE MODELING PROCESS

UNCONDITIONAL GRAPH/BRAND PROXIMITY ALGORITHMS

P(conv|φ_{bi})

CONDITONAL PROPENSITY MODEL

P(conv|imp, φ_{bi})

HADOOP DATA STORE

CONVERSIONS

RAW (GRAPH) DATA

Brand Actor Seeds

AD SERVIN
TWO-STAGE MODELING

1. STAGE: UNCONDITIONAL BRAND PROXIMITY & DIMENSIONALITY REDUCTION
   - Goal: estimate the browser-specific brand affinity
   - Draw from some network inference and relational learning techniques to incorporate the signal in the bi-partite graph
   - Can be build prior to campaign start

2. STAGE: CONDITIONAL MODEL
   - Goal: find the best prospect given the opportunity of an ad impression
   - Can potentially incorporate instance-specific features such as time of day, etc.
   - Can be continuously optimizes during the life of campaign
MEASURING BRAND PROXIMITY

BRAND PROXIMITY = \( P(\text{conv}|\varphi_{bi}) \)


Every Impression/Conversion is logged and can be used to train a linear model that combines all graph-based features to maximize probability of a future conversion.

TYPICAL CAMPAIGN ~ 100-10000 PURCHASE CONVERSIONS.
GOAL IS TO PREDICT P(CONVERSION| IMP, φbi)

- For each browser bi, a feature vector φbi can be composed of the various brand proximity measures and impression information.
- The different evidence can be combined via a ranking function f(φbi).

MULTIVARIATE LOGISTIC FUNCTION, TRAINED VIA STANDARD LOGISTIC REGRESSION

- Take full sample of positives, down sample conversions, can build a robust model with <100k examples*
- Using Greedy Stepwise Forward Selection with 10-Fold Cross Validation, can automate the model building process and support >100 custom client models / week.
‘LIFE’ PERFORMANCE OF M6D TARGETING
PERFORMANCE IMPROVES SUBSTANTIALLY WITH MORE DATA AND FEEDBACK LOOP

Month-by-month performance for one large client

Show lift for a particular size targeted population (%)
Left-to-right decreases targeting threshold
The orange horizontal line intersects the curves at the % of the population we can reach to get a 2x lift.

The maroon vertical line intersects the curves at the lift multiple for the best 10% of each population.
<table>
<thead>
<tr>
<th>TARGETING TYPE</th>
<th>IMPRESSIONS</th>
<th>CONV RATE</th>
<th>CPM</th>
<th>ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Retargeting</td>
<td>1,254,094</td>
<td>0.29%</td>
<td>$5.40</td>
<td>11.0x</td>
</tr>
<tr>
<td>Ad Network</td>
<td>626,947</td>
<td>0.13%</td>
<td>$5.00</td>
<td>17.8x</td>
</tr>
<tr>
<td>Ad Network</td>
<td>467,842</td>
<td>0.07%</td>
<td>$5.00</td>
<td>9.6x</td>
</tr>
<tr>
<td>DSP</td>
<td>78,809</td>
<td>0.01%</td>
<td>$5.00</td>
<td>12.7x</td>
</tr>
<tr>
<td>Booking Site</td>
<td>101,970</td>
<td>0.02%</td>
<td>$5.00</td>
<td>16.8x</td>
</tr>
<tr>
<td>Publisher</td>
<td>96,162</td>
<td>0.02%</td>
<td>$5.00</td>
<td>17.1x</td>
</tr>
<tr>
<td>Booking Site</td>
<td>997,471</td>
<td>0.03%</td>
<td>$4.25</td>
<td>5.0x</td>
</tr>
<tr>
<td>Publisher</td>
<td>17,935</td>
<td>0.01%</td>
<td>$10.00</td>
<td>3.5x</td>
</tr>
<tr>
<td>Retargeting</td>
<td>47,851</td>
<td>0.00%</td>
<td>$6.00</td>
<td>1.4x</td>
</tr>
<tr>
<td>Publisher</td>
<td>6490</td>
<td>0.31%</td>
<td>$17.12</td>
<td>2.9x</td>
</tr>
<tr>
<td>DSP</td>
<td>14,270</td>
<td>0.02%</td>
<td>$18.00</td>
<td>14.2x</td>
</tr>
<tr>
<td>Publisher</td>
<td>14,270</td>
<td>0.01%</td>
<td>$18.00</td>
<td>14.2x</td>
</tr>
</tbody>
</table>

M6D CONSISTENTLY DELIVERS SUPERIOR ROI
SO WHY DOES IT WORK SO WELL?

INTEGRATION WITH THE BRAND AND OBSERVING MEANINGFUL BRAND ACTION IS KEY!

• Use supervised learning instead of buying segments
• It might be tempting to consider modeling based on clicks
• No hassle dropping a pixel …
• There are some brands who have no real good online presence

WE FIND GREAT SIGNAL IN OUR BREADTH OF DATA

• Social principles work even with ‘pseudo’ social data
• Having ‘shallow’ long-tail content on wide universe is very useful
• Unique browser pool since we don’t buy or sell
BRAND ACTIONS: CLICKS, CONVERSIONS AND SITE VISITS

- Purchase conversions are rare
- Clicks and Site Visits are common

- Clicks and Purchases are basically uncorrelated!
- Site visits are much better indicators of purchase conversion
WHAT NOT TO DO: USE CLICKS AS A SURROGATE FOR CONVERSIONS
BUT SITE VISITS ARE A GREAT SURROGATE
OFTEN, SITE VISITS IS BETTER THAN CONversions FOR PREDICTING CONversions WHEN THERE ARE RELATIVELY FEW CONVERSIONS

BUBBLE SIZE REPRESENTS #SV/#PURCHASERS

Normalized difference in AUC

PURCHASE BETTER

SV BETTER

Number of Purchasers - Log Scale
"PRIVACY" ONLINE?

Where would we like firms to operate on the spectrum between the two unacceptable extremes:

"You can’t do anything with MY data!"

"We can do whatever we want with whatever data we can get our hands on."

Are there points between the extremes that give us acceptable tradeoffs between “privacy” and efficacy?

ML PROVIDES MANY POSSIBILITIES.
While our targeting is based on social science, we are not collecting a true social network.

There is NO link to the ‘real you’
- What we ‘know’ about you has no meaning in reality
- Double-anonymization

We collect a very thin but wide layer of information

We are part of the IAB: Interactive Advertising Bureau, the driving force of self-regulation

We fall into the category of ‘third party cookies’

You can opt out (see our homepage)

We do not sell data or derivatives
Many exciting opportunities for analytics in the space of advertisement

Machine learning can be the basis for **EFFECTIVE, PRIVACY-FRIENDLY** targeting in online advertising

Important to consider carefully the target used for training models

Predicting Conversions for Targeting On-Line Advertising Using Social Variables Constructed from User-Generated Content - working white paper for Marketing Science Journal.

(Privacy Friendly!) Social Network Targeting for Online Advertising – Invited talk at 2010 SigKDD, presented by Foster Provost.