Targeting

Hari Sundaram
Associate Professor (CS, ADV)
hs1@illinois.edu

thanks: Andrei Broder, Vanja Josifovski

Introduction

Web search

Game Theory

Auctions

Data flows

Privacy

Text Ads

Display Ads

Recommender systems

Behavioral targeting

Emerging areas

Final Presentations

Key components

- Publishers
  - Supply Side platforms (SSP)
- Ad networks, Ad Exchanges
- Demand side platforms (DSP)
  - Advertisers

Companies that aggregate supply from multiple publishers or other intermediaries and matches it with advertiser demand

Often sells inventory not sold as Guaranteed Delivery ("remnant") or from small publishers

Predate exchanges

Ad Networks

Translate from publisher audience (people that visit section X) to advertiser audience (people interested in Y)

- Typically groups ad inventory by categories or demographics
- Create segments of users that cut across multiple sites
- Key proposition that justifies the existence of many of the intermediaries

Margin revenue model: a percentage of transactions

Top 20 Ad Networks/Buy Side Networks, Ranked by US Unique Visitors, June 2017

millions and % reach

	Unique visitors	% reach
1. Google Ad Network	202.7	91.6%
2. Yahoo Audience Network	191.1	86.3%
3. Conversant	178.0	80.4%
4. RadiumOne	165.1	74.6%
5. RhythmOne	164.1	74.1%
6. Exponential	152.4	68.8%
7. Media.net Contextual Ads	151.4	68.4%
8. Gamut—Network	139.9	63.2%
9. Rocket Fuel	131.1	59.2%
10. Amobee	97.5	44.0%
11. Viant	77.9	35.2%
12. AcuityAds	64.0	28.9%
13. Xaxis Publisher Network	58.4	26.4%
14. GumGum	45.6	20.6%
15. AdBlade Network	45.0	20.3%
16. Undertone Networks	39.6	17.9%
17. Bidtellect	36.8	16.6%
18. Collective Display	29.9	13.5%
19. AdSupply	29.0	13.1%
20. Playwire Media	28.3	12.8%

Note: home and work locations; desktop only
Source: comScore Media Metrix Multi-Platform as cited in press release, July 21, 2017

229161 www.eMarketer.com

- Marketplace for trading impressions between ad networks and some large advertisers or agencies
  - Similar to stock exchanges
  - Increase the liquidity of the marketplace by aggregating supply and demand
  - Use auction models to sell and charge for impressions

Ad Exchanges

- Charge per transaction (not percentage)
- Buying possibilities:
  - Bulk buying–conditionally buy multiple impressions at once (e.g. offer $10 CPM for any mid age female on an entertainment page, total budget $1000/day)
  - Real time bidding (RTB)–buy single impression (spot market) – e.g. offer 1c for a particular impression and a particular user (**Not every exchange allows RTB**)

Ad exchanges complicated to use:
- Which exchange?
- When to bid?
- What to bid?

Demand Side Platforms (DSP)

DSP: technology driven optimization for the demand providers
- Unified access to multiple exchanges
- Integration with the data providers to allow Real Time Bidding (RTB)
- Bid and budget management and optimization
- Analysis of results

Trading desks are essentially DSPs but serve a single agency → no competitive pressures/data leaks

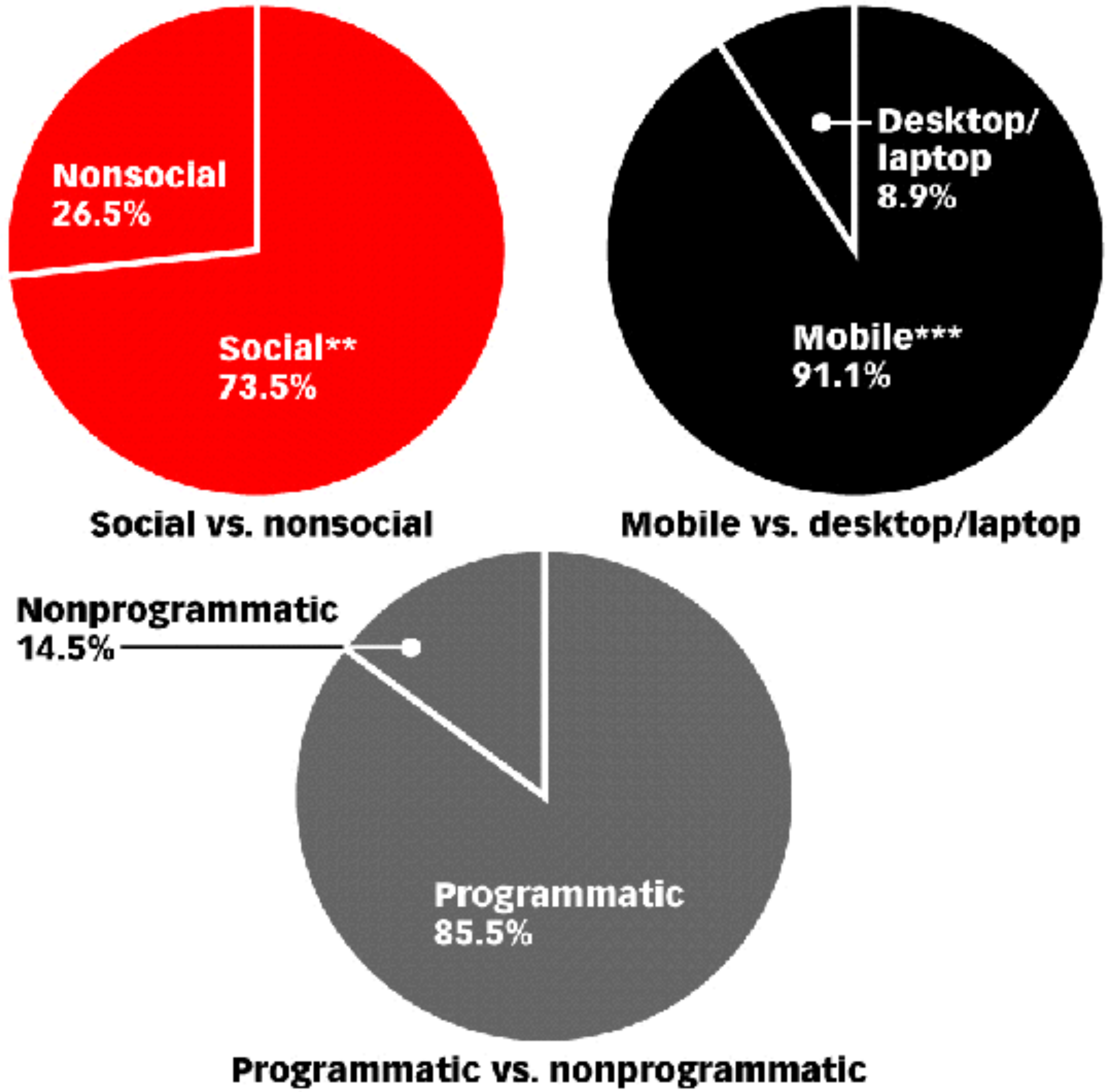
Component of an exchange protocol that allows buyers to bid for each ad impression (context + user) separately in real time

Real-time bidding

The buyers use their own data and targeting options + purchased data to decide how much to bid

RTB bid is usually CPM + **first** price.

US Native* Digital Display Ad Spending Share, by Type, 2018
% of total



*Note: *includes digital display ads that follow the form, feel and function of the content of the media on which they appear; **includes native digital display ads appearing within social networks, social network games and social network apps; ***includes native mobile phone and tablet display ads*
Source: eMarketer, March 2018

235948 www.eMarketer.com

data sale by publishers

GDPR

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 completely 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

re-targeting

Special case of behavioral targeting (more recent, more specific)

User profile generation: describe the user

three tasks

Audience selection:
find the best audience
for a given ad

Performance prediction: find the best ad for a given impression (we have discussed this before)

Understanding the user based on all available data:

- Registration, online activity, offline activity (supermarkets, panels, etc.)

Three basic data item types:

- **Meta data:** name, zip, income bracket, profile interests
- **Activities (events) with a timestamp:** purchases, searches, page views, clicks, …
- **Connections**: friends and others

profile generation

Effective and efficient profile to support the targeting tasks

- **Effective**–right granularity for best modeling performance
- **Efficient**–from Terabytes of data accumulated daily, get a succinct representation

Profile generation: user data → intents and interests

Audience selection and user profile generation are related tasks

- However distinct objectives
- Can be done by different companies

audience selection

Given a campaign find the right audience

Challenges:

- Define the objective
- Training data
- Modeling methods

Audience selection in relation to ad selection ad selection ⟷ audience selection

ad = search (context, user)

user = search (context, ad)

Demographic targeting

traditional targeting

Geo targeting

Behavioral targeting

Retargeting

Important indicator of people’s interest and potential of a conversion

- Imagine you want to sell a $100k sports car. Who do you target?

demographics

Demographic targeting is widely used in traditional media

A must for any display or textual ad network

Widely used in traditional advertising:

- TV, magazines, etc. maintain very detailed statistics of their audience

Common classic dimensions:

- Age
- Gender
- Income bracket
- Location
- Interests (“Golf enthusiast”)

Each dimension has multiple values

User supplied demographic information

- Most reliable—if filled correctly
- In some cases 15-20% of users born on 1st 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.

challenge: where do we get data?

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 information online
- Wider reach—virtually every user

Classification: e.g. regression model on the top of features extracted from history

inference

How to infer the demographics from past behavior?

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

Previous searches/search sessions
Previous browsing activity
Previous ad-clicks
Previous conversions
Declared demographics data Etc.

Behavioral targeting

A technique used by publishers and advertisers to increase campaign effectiveness based on a given user's historical behavior:

Utility—everyone wins! (in principle!)

- **Advertisers**: get a more appropriate and receptive audience, increased conversion rate, better ROI
- **Publishers**: can ask for a premium
- **Users**: see more interesting ads

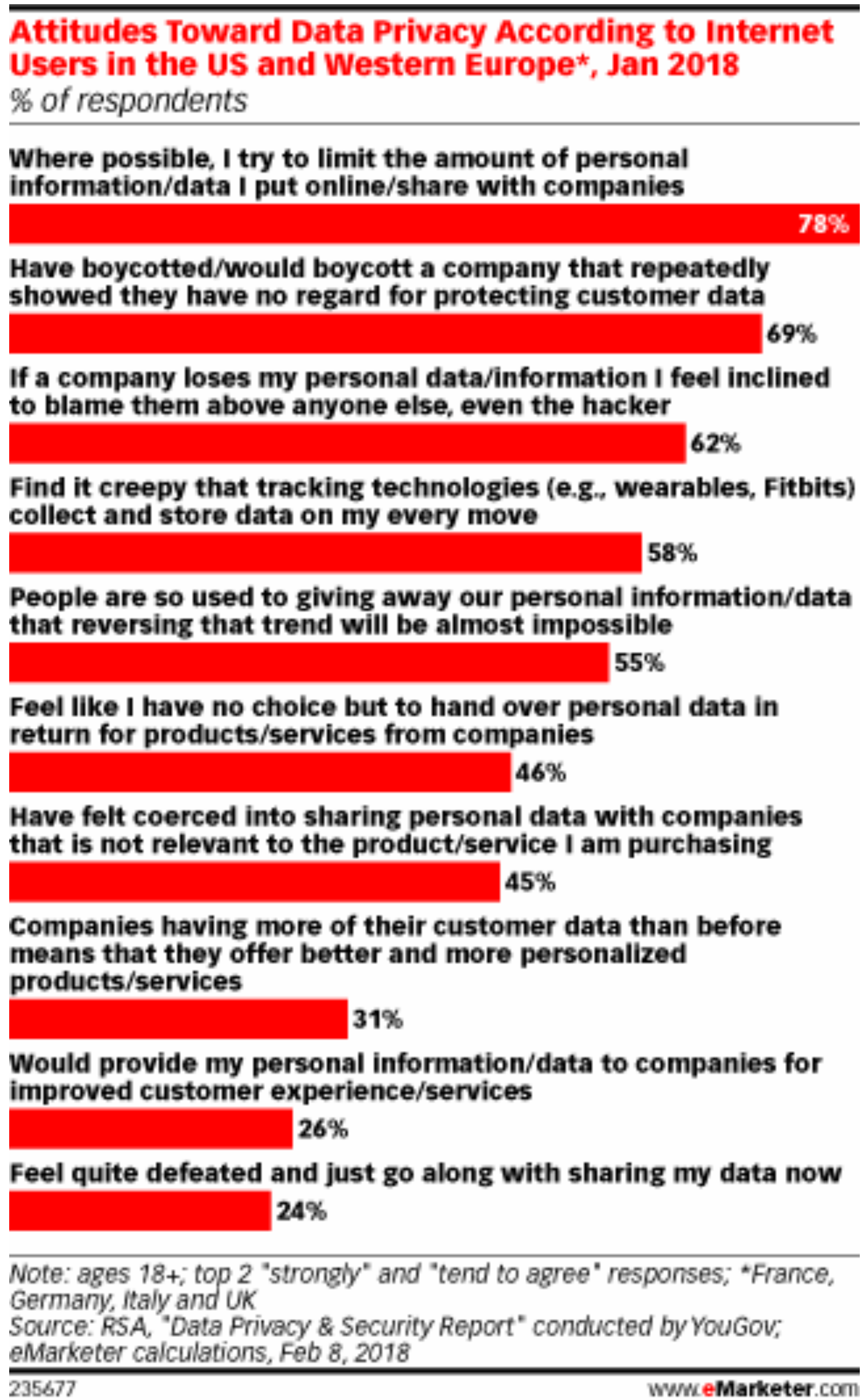
re-targeting

more detailed than other methods

The product you searched
The query you issued
Items put in the cart

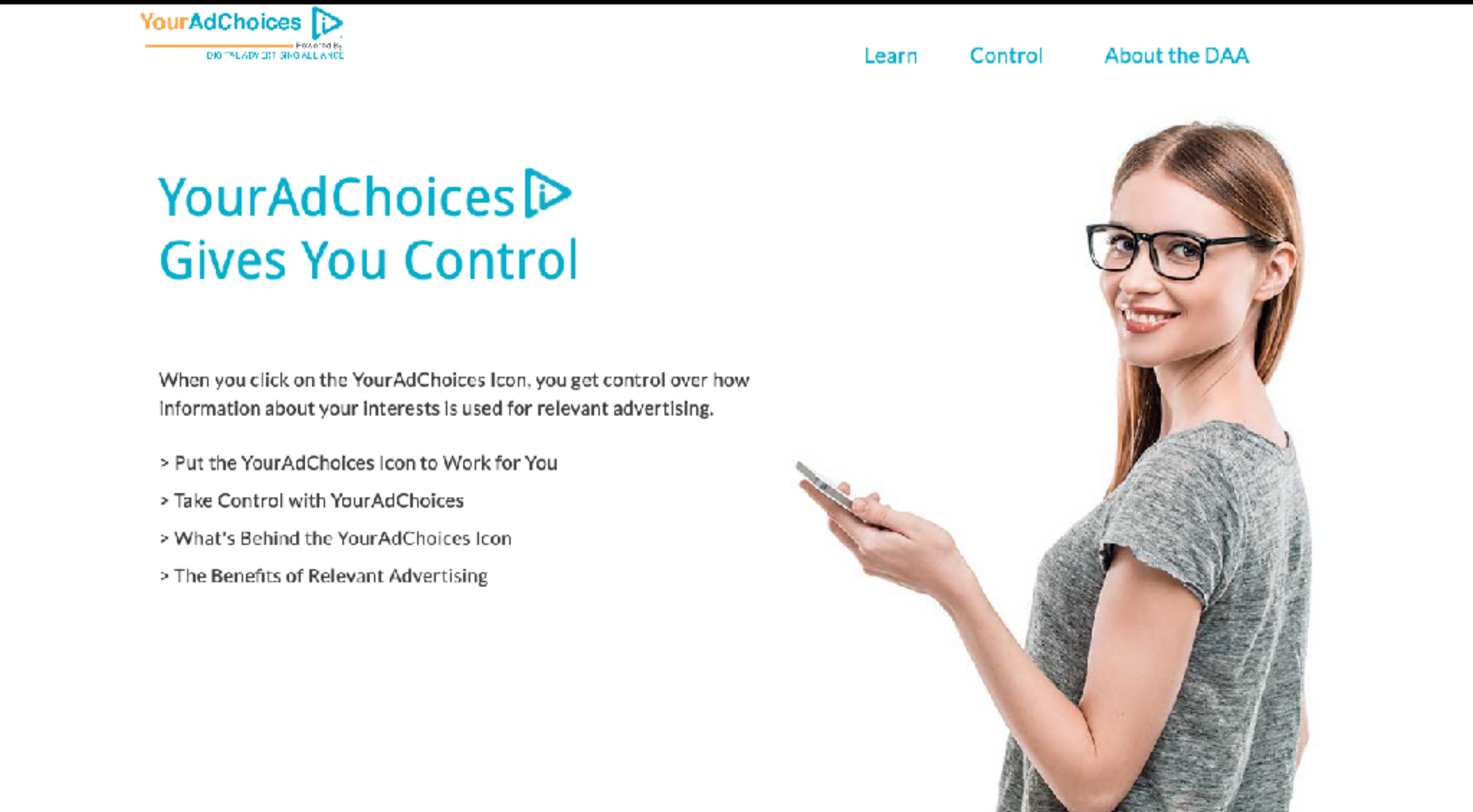
creepy?

does privacy matter?



Attitudes Toward Data Privacy According to Internet Users in the US and Western Europe*, Jan 2018
% of respondents
Where possible, I try to limit the amount of personal information/data I put online/share with companies
78%
Have boycotted/would boycott a company that repeatedly showed they have no regard for protecting customer data
69%
If a company loses my personal data/information I feel inclined to blame them above anyone else, even the hacker
62%
Find it creepy that tracking technologies (e.g., wearables, Fitbits) collect and store data on my every move
58%
People are so used to giving away our personal information/data that reversing that trend will be almost impossible
55%
Feel like I have no choice but to hand over personal data in return for products/services from companies
46%
Have felt coerced into sharing personal data with companies that is not relevant to the product/service I am purchasing
45%
Companies having more of their customer data than before means that they offer better and more personalized products/services
31%
Would provide my personal information/data to companies for improved customer experience/services
26%
Feel quite defeated and just go along with sharing my data now
24%
Note: ages 18+; top 2 "strongly" and "tend to agree" responses; *France, Germany, Italy and UK
Source: RSA, "Data Privacy & Security Report" conducted by YouGov; eMarketer calculations, Feb 8, 2018
235677
www.eMarketer.com

advertising ecosystem solution



Liu, K. and Tang, L. (2011). Large-scale behavioral targeting with a social twist. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, pages 1815–1824, New York, NY, USA. ACM.

Two types of user information
- Profiles attributes
- Social Graph

social targeting

Social networks are the prime destination for display advertising

A mixture of display advertising and textual advertising

Seems like most of the advertising is targeted by explicit selection of 'likes' and other user attributes
- Very precise demographics: e.g. get Chevy owners

Challenges:

- Dilutes the profile information
- How to differentiate different connections?
- How far do you need to go to get all the benefit?

the value proposition

The social graph can be used to smooth the data available about the user

- If you are friends with many people that mountain bike, you are likely to do so as well

Friends are similar along a variety of dimensions is a long- observed empirical regularity (homophily), but:

- Can we leverage friends' activities for behavioral targeting?
- Are forecasts derived from the social graph **more accurate** than standard behavioral targeting models?

the IM social graph

Investigated a wide array of supervised and unsupervised machine-learning approaches to utilize social data for BT models

Evaluated the predictive power of social data across 60 consumer domains on a network of over 180 million users in a period of 2.5 months.

BT qualifications

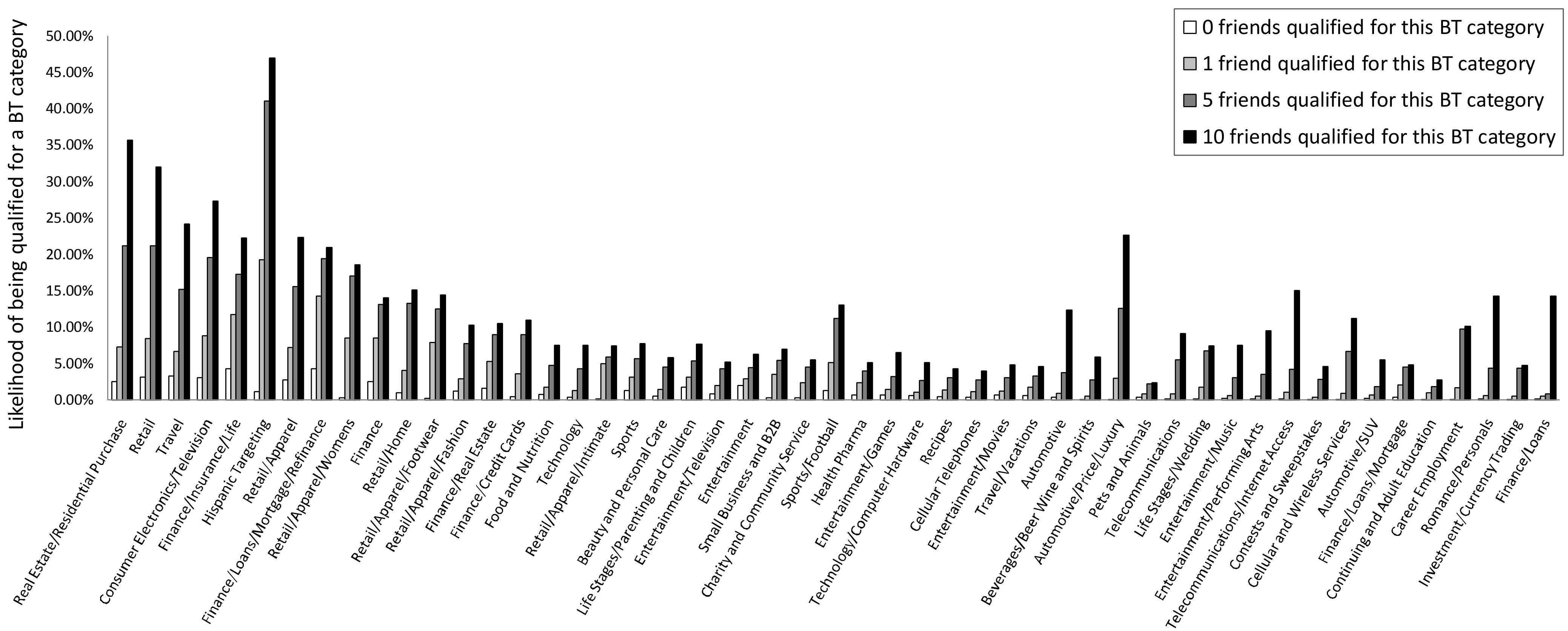


Figure 1: Likelihood of being qualified for a BT category as a function of having social contacts who are also qualified for the same category.

Homophily—ad clicks

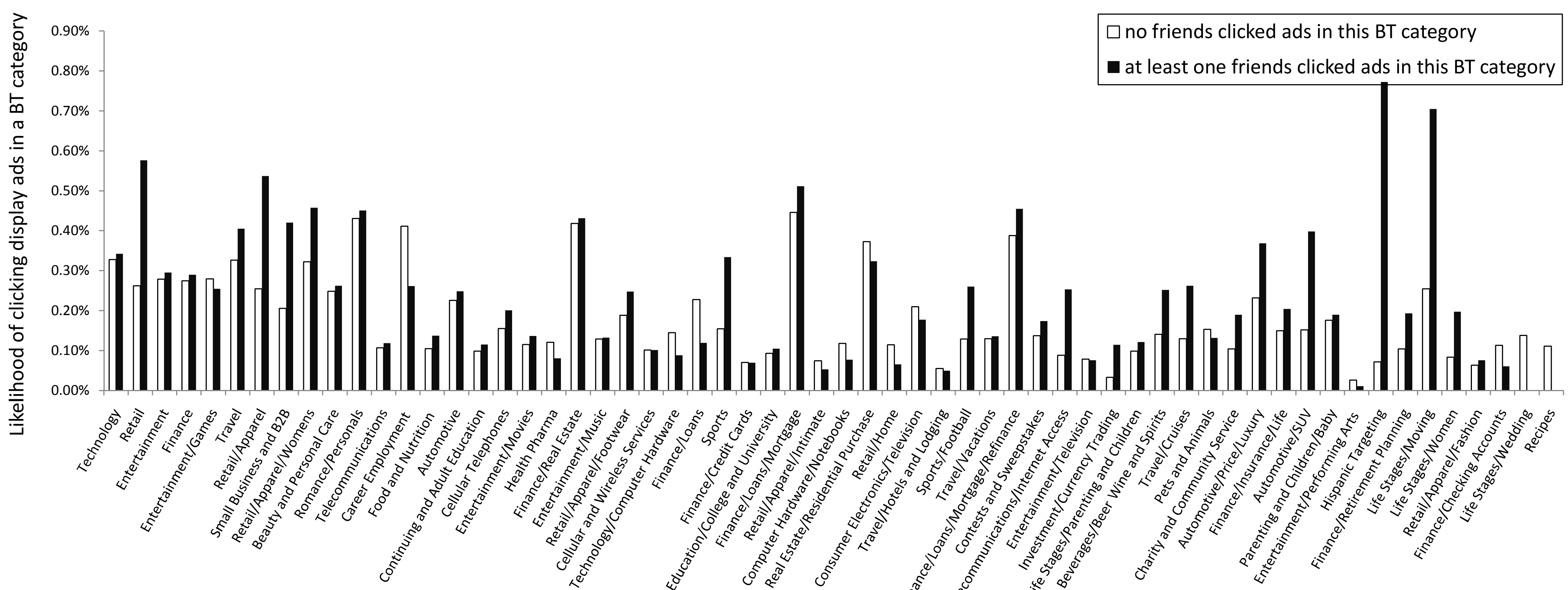


Figure 2: Likelihood of clicking on display advertisements in a BT category as a function of having social contacts who have done the same before.

Classification: append social features to standard BT features, then train models (supervised approach)

Leveraging social data

Ensemble: combine BT model and social model (supervised approach)

Network propagation: use propagation to infer BT scores from one’s neighbors (mostly unsupervised approach)

experiments

	Method	$\Delta_{\overline{\text{AUC}}}$
1	**BT baseline**	0.00%
2	**Random targeting**	−24.25%
3	**Neighbors1**	−16.58%
4	**Neighbors2**	−13.34%
5	**Community**	−15.84%
6	**BT + Neighbors1**	+0.97%
7	**BT + Neighbors2**	+0.86%
8	**BT + Community**	+0.08%
9	**Ensemble**	+0.00%

category level improvement

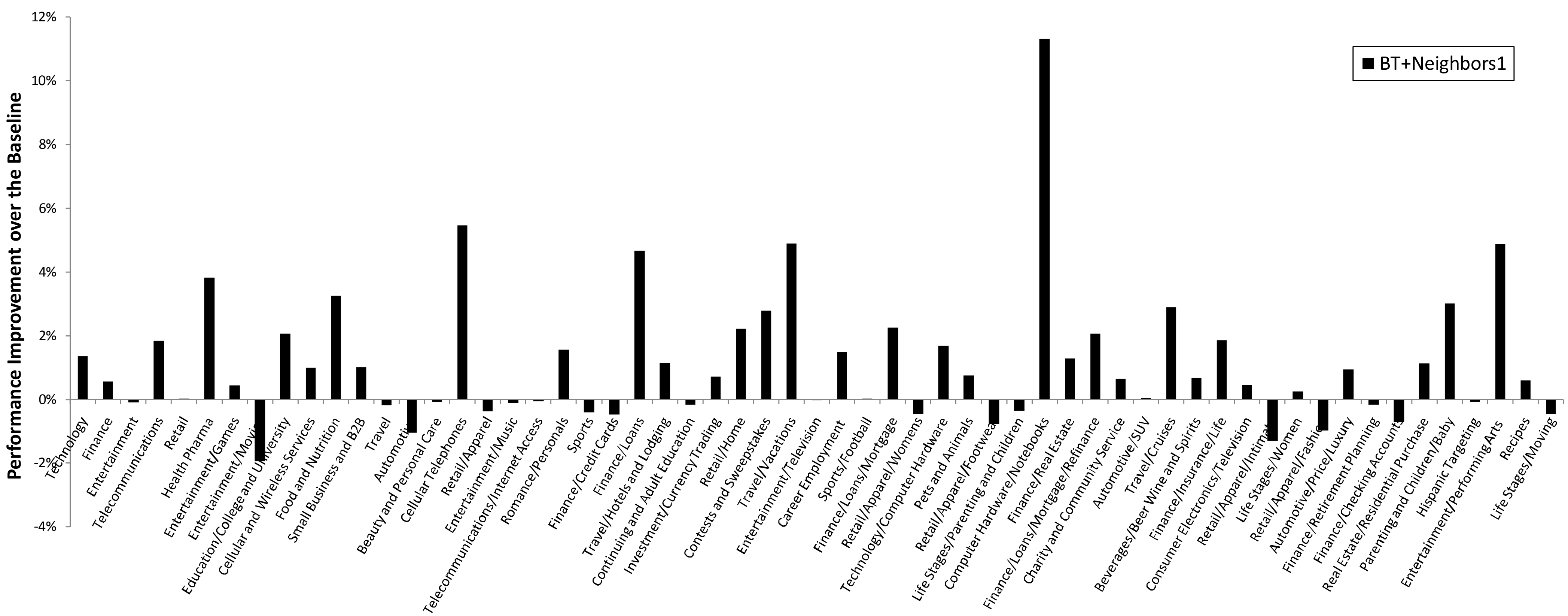


Figure 7: Category-level performance improvement, measured by the relative lift of AUC, *i.e.*, $(\mathbf{AUC}_c/\mathbf{AUC}_{\mathbf{baseline}} - 1) \times 100\%$. Categories are sorted in descending order of ad impressions served in that category. The method is BT + Neighbors1.

BT baseline model **substantially outperforms** all other models trained from social features only; indicating that individual's own behavioral is much more useful than their friends.

takeaways

Social features do carry informative signals.

Note this study uses the IM graph
Implicit feedback from the user

In social networks the graph induces different behavior explicit and implicit feedback

Personas share interests and behavior

Usually pre-defined segments based on real world behavior

Same user can have multiple personas: **personas are facets of personality**

not just demographics

Personas change over time

advertisers want to target personas

happy Mom

busy executive

entrepreneur

Personal vs. global

broken microwave elections

General vs. specific (sports vs. Inter Milan)

Short term (current) vs. long term interest

Alfredo Rodrigez Jazz

user interest dichotomy

Distinguishing between these would allow for more effective advertising

Online tasks have limited time span

What is of interest now, might not be in a few minutes/ hours/days:

- Buy a cruise, don't need another one for a while

short-term v. long term

Other interest persist

- Tennis, Roger Federer

What is better suited for advertising?

Can we decide on the generality vs. specificity of an interest of a user?

Specific interest can be targeted much better

- Requires adequate ad supply

General interest has better coverage in the ad supply

general v. specific

What is the relationship between general-specific and short-long term interests?

- Short term interest more specific (Lady Gaga)
- Long term interest more general (pop music)

Help differentiate short term interests by first understanding the long term interests and use those as prior in the interpretation of recent events

profiling with language models

Tyler, S. K., Pandey, S., Gabrilovich, E., and Josifovski, V. (2011). Retrieval models for audience selection in display advertising. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, pages 593–598, New York, NY, USA. ACM.

Given: seed set of users that have performed a desired action (e.g. bought a product) for a campaign

Find: more users that are likely to do so.

intuition

Finding such users can be formulated as a retrieval task:

- Top-**k** similarity search over the user space
- Query is composed from the seed set

ad clicks, page views etc.

$$u = \{e_1, e_2, \ldots, e_n\}$$

events

problem definition

the **goal**: learn to distinguish convertors from non-convertors

non-converters are not necessarily negative examples– many reasons why a person will not click on an ad

IR: model of the queries and documents, and their similarity

retrieval models

Successfully applied in IR

- Vector space
- Language models

Can we use **similar tools** to formalize the audience selection task in display advertising targeting?

- Users instead of documents
- Query derived from the seed set

$$u = \{e_1, e_2, \ldots, e_n\}$$

$$e_i = < \text{type}_i, \text{int}_i, c_i >$$

language model

generate the user from an underlying distribution

$$\begin{aligned} p(u) &= p(e_1, e_2, \ldots, e_n) \\ &= p(e_1)p(e_2 \mid e_1)p(e_3 \mid e_2 e_1)\ldots p(e_n \mid e_1 e_2 \ldots e_{n-1}) \\ &\sim \prod_{i \in 1\ldots n} p(e_i) \end{aligned}$$

assuming independence

generating an event

$$p(u) \sim \prod_{i \in 1\ldots n} p(e_i) \sim \prod_{i \in 1\ldots n} p(\text{int}_i) \cdot p(\text{type}_i \mid \text{int}_i) \cdot p(c_i \mid \text{int}_i, \text{type}_i)$$

prob of observing an event in a given time interval; assume it to be proportional to the length of the interval

prob of observing in that **interval** an event of the given type; assume that the event mix is independent of the interval

prob of observing a specific event **content** given the interval and the event type

$$p(c_i \mid \text{int}_i, \text{type}_i) = \prod_{w \in c_i} p(w \mid \text{int}_i, \text{type}_i)$$

seed set representation

What would be a good representation of the seed set?

- A language model of the seed set –probability distribution
- Close to the seed set
- Far from the background

seed set representation

background model composed of all other users

$$p(w \mid \theta^q) = \exp\left(\frac{1}{1-\mu} \frac{1}{n} \sum_{i=1}^{n} \log p(w \mid \theta^i) - \frac{1}{1-\mu} p(w \mid C) \right)$$

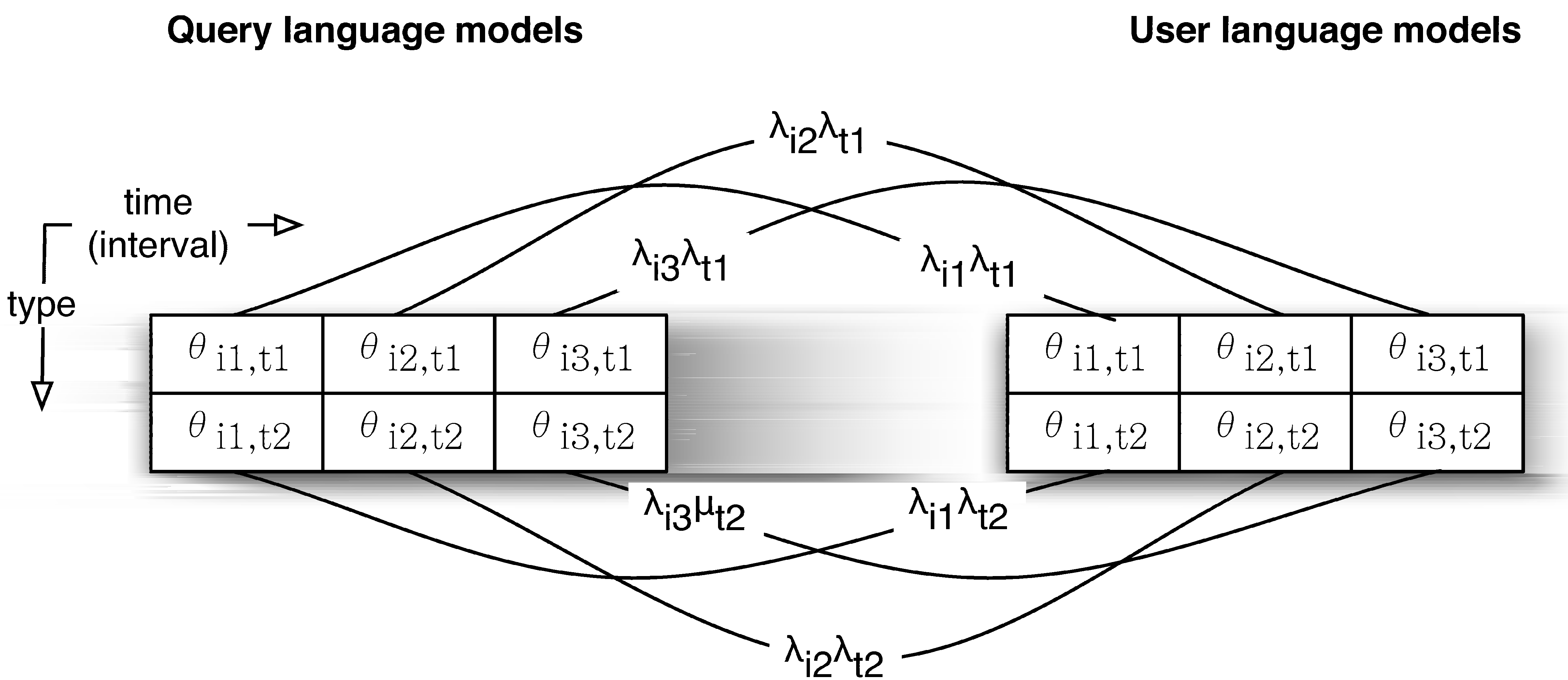
parameter
learned on a
validation set

the models for the users in the seed set

does time matter?

So far, there is no time aspect to our model–
how to introduce time?

- Ignore time
- Sequence of events–each event in its own time interval (as in Markov processes)
- Something in between–intervals



temporal models

Dataset: 34 different ad campaigns from Yahoo advertising network

Seed set: users who converted on these campaigns between 02/04/2010 and 02/18/2010

Test set: ad impressions between 02/19/2010 and 02/24/2010

Each user profile has 4 weeks of her online activity prior to the ad impression

experiments

"We found that our approach performs well in all three campaign types, in fact it does better on the medium (AUC-ROC: 0.657) and small (0.673) campaigns compared to the large ones (0.624)."

Time intervals to capture the temporal aspect at good granularity

Compare the seed set activity within the same window of the target users

Users with similar activity at the similar interval with the seed set are ranked higher

retrieval models

Provide solid formalization for audience selection

Can use multiple models
- Language modeling
- Vector space model

- Traditional targeting: demographic, geographic, behavioral
  - How to get the data from the user?
  - Infer the data from historical activity

summary

Targeting is a key step in differentiation of impressions and extracting value

- User profile generation is key:
  - Generative models to assign probability of a sequence of events
  - Weighting based on time, event type and content
  - Predict the counts of events in certain categories

Introduction

Web search

Game Theory

Auctions

Data flows

Privacy

Text Ads

Display Ads

Recommender systems

Behavioral targeting

Emerging areas

Final Presentations