What do they know about me? **Contents and Concerns of Online Behavioral Profiles**

Ashwini Rao, Florian Schaub and Norman M. Sadeh
arao@cmu.edu, fschaub@cs.cmu.edu, sadeh@cs.cmu.edu

### Background

Online services landscape driven by data economy
- data aggregators, service providers, advertisers etc.
- collect and trade user data from multiple sources
- create behavioral profiles

![Sample behavioral profile](image)

### Pros and Cons of Data Economy

**Benefits**
- Free services supported by advertising
- Risk mitigation e.g. id verification
- People search
- Consumer insights

but **privacy concerns** about behavioral advertising\(^2\), people search\(^3\) etc.

### Problem Statement

To understand tradeoffs better, we investigate behavioral profiles of users

- What **types of data** are present in behavioral profiles?
- What **concerns** do users have about the data in their profiles?

### Methodology

**In-person interviews** (n=8)
- Asked participants to **look at their own profiles**
- Elicited concerns and identified data types
- Studied BlueKai Registry, Google Ad Settings and/or Yahoo Ad Interests profiles

**Online survey** (n=100)
- Validated identified concerns with a larger group

### Some Data Types we found in Behavioral Profiles

<table>
<thead>
<tr>
<th>Geographic</th>
<th>US &gt; Massachusetts &gt; Boston-Cambridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>IP address – 71.182.182.9</td>
</tr>
<tr>
<td>Demographic</td>
<td>Female, single, 20-24 years Some college, IT professional Income Range – $75-$90K Household Size – 1, Number of Adults – 1 Children in Residence – No Home Value – 0-$100,000 Voter Indicator</td>
</tr>
<tr>
<td>Psychographic</td>
<td>Religion – Value Tiers 1-3 Past purchase &gt; ISP &gt; Internet &gt; Verizon Offline CPG Purchasers &gt; Charmin Ultra Soft CPG &gt; OTC Medicine/Cough and Cold &gt; pain reliever Buy American – Not Likely</td>
</tr>
<tr>
<td>Behavior</td>
<td>Owns a regular Amex card Prepaid Wireless Plan Subscriber Eco Friendly Vehicle Owner</td>
</tr>
<tr>
<td>Predictive</td>
<td>Credit Card Interest Score – 16-17% Credit Card App Intent Score – 10-11% Auto insurance online buyer – High Propensity In-Market – Cell-Phones and Plans</td>
</tr>
<tr>
<td>Life event</td>
<td>Empty nester</td>
</tr>
</tbody>
</table>

### Validating Identified Concerns

Does a broader audience agree with the identified concerns?
- Conducted online survey (n=100) using a sample profile (Fig. 1) on AMT platform\(^4\)

![Survey results](image)

At least **70%** agree with each concern

### Discussion

- Users have several concerns including sensitive data coming from offline sources
- Profiles are highly inaccurate – who is accountable for adverse impact?
- What does it mean to delete data from a profile?
- Comprehending and accessing profile data is difficult – usability studies required

### User Concerns Identified

1. Collection of sensitive data
2. Extent of collection
3. Combining data from many sources
4. Level of detail
5. How data may be used
6. Accuracy of data
7. Impact of editing profile data

### References

2. Ur et al. “Smart, useful, scary, creepy: perceptions of online behavioral advertising,” SOUPS 2012
4. Amazon Mechanical Turk [www.mturk.com](http://www.mturk.com)