Influence in Classification

Anupam Datta

CMU
Fall 2015
Big Data Analysis and Transparency

• Big data is big business.
• It is “good”: able to identify trends, produce accurate results.

It is not transparent!

• It is hard to tell what factors determine classification outcomes.
System Model

Personal data of many users

Big Data System (Classifier)

Decision for individual users

Goal: Measure influence of features on classifier’s decision
Measuring Feature Importance

How important was the $i$-th feature in determining the classifier’s output?
Causal Influence Measure: Idea
[Datta, Datta, Procaccia, Zick 2015]

Counts number of times a change in state causes a change in classifier’s decision.
Notation

• A set of **features** \( N = \{1, \ldots, n\} \)
• For each \( i \in N, A \downarrow i \): set of possible **states**.
• \( A = \prod i \in N \uparrow \forall A \downarrow i \) : all possible profiles.
• \( v : A \rightarrow \{0,1\} \), labels data.
• Dataset: \( \langle B, v \rangle \), where \( B \subseteq A \)  
  (we don’t see all profiles)
Notation

• An influence measure: a function $\varphi$ that, given a dataset $\langle B, v \rangle$, outputs a value $\varphi \downarrow i$ for every feature $i \in N$.

“how important is gender for this classification?”
Causal Influence Measure
[Datta, Datta, Procaccia, Zick 2015]

\[
\psi_{\downarrow i}(\mathcal{B}_v) = C \cdot \chi_{\downarrow i}(\mathcal{B}_v);
\]

and is a constant independent of (but may depend on ).
Relation to Linear Classifiers

**Theorem:** suppose that $v$ is a linear classifier, defined by $w^{\downarrow 1}, \ldots, w^{\downarrow n}$ and $q$. Then $\chi^{\downarrow i}(A, v) \geq \chi^{\downarrow j}(A, v)$ if and only if $|w^{\downarrow i}| \geq |w^{\downarrow j}|$.

High weight translates to high influence!
Implementation

• To test our measure’s behavior, we measure influence on a generated dataset.

• We employ the AdFisher framework [Datta et al. 2014] to create fake Google user profiles and observe the ads that they are presented.
Experimental Setup

• 12 x 100 simulated users, different setting of
  – Gender: male or female
  – Age: 18-24, 35-44, 55-64
  – Language: {English, Spanish}

• Go to bbc.com/news, collect the ads displayed.
Influence of Features

• \( v(\alpha) \) for profile \( \alpha \): a vector measuring counts of for each unique ad served

Example: \( v↓i(M,18-24,En) = \text{# of times i}^{\text{th}} \text{ unique ad was displayed to this profile} \)

• Influence measures for:
  – Gender: 0.124; Age: 0.120; Language: 0.141

• Language most influential but not by much
Future Work

• Leverage knowledge of priors on data
• Account for correlated features
Goal: Measure influence of features on classifier’s decision
Extensions

• **State Influence:** how influential is being 25-35, vs. how important is age.

• **Generalized distance measure:** replacing $\|v(a) - v(a↓−i, b)\|$ with a pseudo-distance $d(a, (a↓−i, b))$. 
# Top Ads for Age

<table>
<thead>
<tr>
<th>Title/Ad Description</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.</td>
<td>0.07</td>
</tr>
<tr>
<td>Jim Rickards Project 2015/Economist, Jim Rickards explains the coming economic crash.</td>
<td>0.0663</td>
</tr>
<tr>
<td>”My Insomnia Trick”/Naturally Fall Asleep Fast, Stay Asleep All Night – Wake Up Refreshed</td>
<td>0.0661</td>
</tr>
<tr>
<td>Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth</td>
<td>0.0611</td>
</tr>
<tr>
<td>Sciatica Exercises?/Stop: What You MUST know Before attempting to Treat your Sciatica:</td>
<td>0.0606</td>
</tr>
</tbody>
</table>
# Top Ads for Gender

<table>
<thead>
<tr>
<th>Title/Ad Description</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim Rickards Project 2015/Economist, Jim Rickards explains the coming economic crash.</td>
<td>0.07</td>
</tr>
<tr>
<td>Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.</td>
<td>0.0583</td>
</tr>
<tr>
<td>Tech Gadgets/Daily Deals on Modern Gadgets. Exclusive Pricing - Up To 70% Off.</td>
<td>0.0564</td>
</tr>
<tr>
<td>Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth</td>
<td>0.0561</td>
</tr>
<tr>
<td>Elabore su Presupuesto/Nuestros Consejeros Certificados Est´an listos para ayudarlo</td>
<td>0.0534</td>
</tr>
</tbody>
</table>
## Top Ads for Language

<table>
<thead>
<tr>
<th>Title/Ad Description</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elabore su Presupuesto/Nuestros Consejeros Certificados Est´an listos para ayudarlo</td>
<td>0.1667</td>
</tr>
<tr>
<td>The Greatest Penny Stocks/Get free daily penny stock alerts. Join now. New pick out soon.</td>
<td>0.0755</td>
</tr>
<tr>
<td>Business Leads CRM/Business Lead Manager, Dialer, CRM. 400% Boost in Conversion Rates.</td>
<td>0.0683</td>
</tr>
<tr>
<td>Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth</td>
<td>0.0644</td>
</tr>
<tr>
<td>Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.</td>
<td>0.06</td>
</tr>
</tbody>
</table>