Practical and Ethical Considerations in Demographic and Psychographic Analysis

Jason Baldridge
Co-founder & Chief Scientist
People Pattern
@jasonbaldridge
Brief background

Spanning academia and industry
People > Posts

(But language provides a remarkable window into people and communities.)
Conversation Focus

I love coffee!!!
Conversation Focus

Authors

John Smith
cool42@example.com
Atlanta, GA, USA

Texts

I love coffee!!!
John Smith
cool42@example.com
Atlanta, GA, USA

I love coffee!!!
The Problem

Identify, segment and analyze groups of people.
Identification

keywords, hashtags, demographics, stitching
Identification

keywords, hashtags, demographics, stitching
Segment and Analyze

profiles, posts, images, connections, clustering
Segment and Analyze

profiles, posts, images, connections, clustering
**Tailored audiences**

<table>
<thead>
<tr>
<th>Shopping (29%)</th>
<th>Eating (24%)</th>
<th>Family (20%)</th>
<th>School (17%)</th>
<th>Music (10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#giveaway</td>
<td>#ff</td>
<td>#parenting</td>
<td>#kids</td>
<td>#soundcloud</td>
</tr>
<tr>
<td>#win</td>
<td>recipe</td>
<td>child</td>
<td>student</td>
<td>video</td>
</tr>
<tr>
<td>#health</td>
<td>amazing</td>
<td>baby</td>
<td>public</td>
<td>liked</td>
</tr>
<tr>
<td>#ad</td>
<td>#health</td>
<td>birthday</td>
<td>elementary</td>
<td>added</td>
</tr>
<tr>
<td>#quote</td>
<td>morning</td>
<td>dad</td>
<td>career</td>
<td>warhol</td>
</tr>
<tr>
<td>new</td>
<td>wait</td>
<td>adhd</td>
<td>speech</td>
<td>church</td>
</tr>
<tr>
<td>free</td>
<td>tonight</td>
<td>sick</td>
<td>fresher</td>
<td>playlist</td>
</tr>
<tr>
<td>happy</td>
<td>loss</td>
<td>smile</td>
<td>charter</td>
<td>🎵</td>
</tr>
<tr>
<td>want</td>
<td>body</td>
<td>thankful</td>
<td>sunset</td>
<td>whitney</td>
</tr>
</tbody>
</table>

Interest prediction and extraction of interest-specific keywords. Promoted tweet copy informed by persona-based keywords.
Lessons from industry

The needs of industry expose interesting challenges!
Complete coverage

Users want all the attributes for all the profiles, but mistakes will be made.
Cherries & cockroaches

One cockroach spoils the bowl!
Daniel Kahneman, “Thinking, Fast and Slow”
High precision

Analyze model confidence and precision/recall tradeoff.
Users need aggregate statistics for arbitrary segments.
Segment & aggregate

High precision thresholding leads to different recall proportions for different classes and changes aggregate statistics.
Soft aggregations

Store and use both high precision labels and classifier confidence distributions.
Age granularity

Supporting year-level granularity supports multiple use cases.
Age granularity

Supporting year-level granularity supports multiple use cases.
Supporting year-level granularity supports multiple use cases.
Supporting year-level granularity supports multiple use cases.
Supporting year-level granularity supports multiple use cases.
Age granularity

Supporting year-level granularity supports multiple use cases.
Age granularity

Supporting year-level granularity supports multiple use cases.
Age granularity

Supporting year-level granularity supports multiple use cases.
Supporting year-level granularity supports multiple use cases.
Regulatory requirements

21 or older: ≥85%

Aggregate accuracy matters.
Regulatory requirements

21 or older: $\geq 85\%$

Aggregate accuracy matters.
Regulatory requirements

21 or older: ≥85%

Aggregate accuracy matters.
Actual or apparent age?

59 / 57  37 / 35  65 / 51  20 / 29

Actual / apparent (age)

Looking At People 2015 Apparent Age Challenge:
Labels are 10 or more age guesses by people given photo.

Use actual age (if you can)!

Actual age matters for many tasks. Annotating a person’s birth year is also future proof.
Use actual age (if you can)!

Actual age matters for many tasks. Annotating a person’s birth year is also future proof.

```json
{
    "target_uuid": "twitter:119837224",
    "label": "1974",
    "namespace": "birthyear",
    "context": {
        "creator": "John Smith",
        "creator_type": "human",
        "confidence": 4,
        "date": "2016-07-18T14:11:05Z",
        "note": "Verified by resume."
    }
}
```
Use actual age (if you can)!

Actual age matters for many tasks. Annotating a person’s birth year is also future proof.
@johnsmith1974

42yo. Dad of 2 amazing girls. Cubs fan. All-around happy guy.

23. Proud Longhorn. Still crazy about Justin Bieber!
Annotation by patterns

https://arxiv.org/abs/1601.04621
Annotation by patterns

"Detecting the age of Twitter users."
https://arxiv.org/abs/1601.04621

"Your age is no secret."
Many high-quality annotations can be extracted, but the resulting age distributions are highly skewed.
Actual or apparent race?

It’s well beyond skin-deep.
http://bcomposes.com/2015/11/08/improving-race-relations-a-path-forward/
Actual or apparent race?

It’s well beyond skin-deep.
http://bcomposes.com/2015/11/08/improving-race-relations-a-path-forward/

It’s complicated.

Sen and Wasow (2016).
“Race as a bundle of sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics”
Actual or apparent race?

There is no “actual”. Race is as a concept is fluid across individuals, communities, and time periods.

http://www.understandingrace.org/

https://twitter.com/zeynep/status/793805186892492800
Complete coverage

We can’t arbitrarily ignore profiles or certain class labels.
Complete coverage

We can’t arbitrarily ignore profiles or certain class labels.
Complete coverage

Not all profiles are people!
Account type classifiers are necessary.
Common featurization

Strong common feature representations facilitate deployment, improvements and tackling new tasks.
Low dimensional features can be combined with a convolutional neural network for profile picture.
Image manipulation

Image classifiers quickly pick up on stereotypes.

https://medium.com/@kerryrodden/is-that-a-boy-or-a-girl-cb93abbae6da
“Finally, this project started out as a fun personal curiosity, but the most important thing I have learned from it is very serious: looking at the misclassifications helped me reflect on the risk that a gender classification model could be misinterpreted and misused. For example, maybe after reading this article, you wonder if it would be fun to create a new app that can automatically rate how “masculine” or “feminine” a person looks in a photo. But then imagine it being used by an insecure teenager, or by a high school bully who applies it to all of the photos in the class yearbook. In most of the world, gender norms are rigid, and come with very strong pressure to conform—and people who do not conform (or are even perceived not to) face harassment and threats to their personal safety.”

- Kerry Roden

Image classifiers quickly pick up on stereotypes.

https://medium.com/@kerryrodden/is-that-a-boy-or-a-girl-cb93abbae6da
Positive use: GD-IQ

Automating the Geena Davis Inclusion Quotient turned a months long process into real time analysis—and uncovered important gender patterns in film-making.

Handle with care

Demographic ad creation and targeting

Predictive policing

Models and algorithms have the potential to reinforce or ameliorate existing social biases
Discrimination pitfalls

ProPublica was able to create a Facebook ad regarding housing that excluded African-Americans & Hispanics.
Discrimination pitfalls

ProPublica was able to create a Facebook ad regarding housing that excluded African-Americans & Hispanics.
ML for anti-discrimination

Maximize profits

Equal opportunity

Use thresholds that attempt to create equality of opportunity across multiple groups, rather than maximizing profits.

ML for anti-discrimination

Setting thresholds requires knowing the protected class values. Fairness in lending, hiring, etc. using social data thus requires such predictions.

Standard (out-of-the-box) NLP tools perform worse on AAE, an example of how current models/methods/methodology could have negative social impact on some groups.

Psychographics

Can we characterize psychographic traits via information contained in profile data (posts, graph, images)?
Dark Tetrad

The favorite activity of people who score highly for the dark tetrad personality types is…. surprise… trolling!

Comparing banned & normal users (in retrospect): banned users wrote posts that are less relevant, harder to read, and less positive.

Cheng et al. (2015). “Antisocial Behavior in Online Discussion Communities.”
The standard approach is to use the Big 5 personality traits.

Features include both content and stylistic observations, often derived from LIWC.

Prediction performance measured against results from standard personality tests.

Image from Adali and Golbeck (2015) “Predicting Personality with Social Behavior”
Personality classification

Bloggers (n = 692)

Language production provides a window on personality at scale.

Yarkoni (2010). “Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers.”
Twitter users whose language indicates higher openness and lower neuroticism are more likely to respond positively to an ad.

Chen et al. (2015). “Making Use of Derived Personality: The Case of Social Media Ad Targeting.”
**McAdams’ three levels of personality**

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Dispositional Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Actor</td>
<td>General tendencies, e.g Big Five traits.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 2</th>
<th>Characteristic adaptations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivated Agent</td>
<td>Beliefs, desires, coping mechanisms.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 3</th>
<th>Narrative Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autobiographical author</td>
<td>Life stories that define meaning and purpose.</td>
</tr>
</tbody>
</table>

Personality is more than a weighted combination over five categories! See also Brian Little’s “Me, Myself and Us” and Jonathan Haidt's “The Happiness Hypothesis.”

McAdams 1995, “What do we know when we know a person?”, [https://www.sesp.northwestern.edu/docs/publications/557464623490a3fc35faeb.pdf](https://www.sesp.northwestern.edu/docs/publications/557464623490a3fc35faeb.pdf)
Moral foundations theory

Care / Harm
Fairness / Cheating
Liberty / Oppression
Loyalty / Betrayal
Authority / Subversion
Sanctity / Degradation

Six fundamental moral foundations

Moral foundations theory

Moral foundations and political leaning.

http://www.yourmorals.org/
Moral foundations

How Candidate Supporters Prioritize Moral Foundations
Compared to the Average American
(Average Responses Standardized)

- Care (Empathy)
- Proportionality (Just Deserts)
- Liberty
- Loyalty-Authority-Sanctity

US 2016 presidential candidates and moral foundation differences.

http://righteousmind.com/presidentialprimaries/
Majority illusion

The connectedness of opinion holders greatly impacts the perception of others. A minority opinion can appear extremely popular for each individual (left side).

Majority illusion

The connectedness of opinion holders greatly impacts the perception of others. A minority opinion can appear extremely popular for each individual (left side).

The connectedness of opinion holders greatly impacts the perception of others. A minority opinion can appear extremely popular for each individual (left side).

Filter bubbles

Algorithms tend to show us more of what we like. Should they also show us things that challenge us?

See my post on the Emotional Contagion study: http://go.peoplepattern.com/blog/emotional-contagion-one/
Measuring Implicit Bias from Quoting Patterns

Personal estimates of bias are unreliable.
Behavior in context reveals consistent, interpretable patterns.

Niculae et al. (2015). "QUOTUS: The Structure of Political Media Coverage as Revealed by Quoting Patterns."
Personality and moral foundations predictions from text are typically based on word counts and/or word count featurization using a curated lexicon (e.g. LIWC and MFD).

The lexicons can be scaled and improved with models and unlabeled data, including to apply them to other languages and dialects.

Deeper analysis could reveal new linguistic features that correlate with personality/morality, including syntactic constructions and discourse structure.

Being able to predict patterns of demographics, personality and morality accurately and at scale could help us combat abusive behavior and better understand and address current social & political divisions.


Aggregate analysis

We should enable better aggregate understanding that doesn’t require individual level annotations and predictions.
Ranked multi-labels

**Income**: medium-high, high, medium

**Race**: white, hispanic, black, east-asian

**Parent**: no, yes

**Political affiliation**: democrat, republican, independent

Attach preference distribution for multiple labels to a topic.
Ranked multi-labels

- **Income**: medium-high, high, medium
- **Race**: white, hispanic, black, east-asian
- **Parent**: no, yes
- **Political affiliation**: democrat, republican, independent

Attach preference distribution for multiple labels to a topic.
Distant Demographics

Labels propagate through both profile and feature nodes.
Labels propagate through both profile and feature nodes.
Distant Demographics

Labels propagate through both profile and feature nodes.
Distant Demographics

Labels propagate through both profile and feature nodes.
Distant Demographics

Labels propagate through both profile and feature nodes.
Distant Demographics

Labels propagate through both profile and feature nodes.
Log-likelihood ratios

#trumpence2016

#maga

#benghazi

#hillary2016

#strongertogether

#blacklivesmatter
Beyond text

Digital social networks may become much more “social”. Language will be both textual and spoken, which will provide many interesting new challenges, opportunities, and dilemmas.

http://www.dailymail.co.uk/sciencetech/article-3523598/Oculus-fire-revealed-VR-headset-s-T-Cs-allow-Facebook-collect-information-physical-movements.html
Our tricky reality

The work we do in NLP is interesting and necessary—all the more so with the massive influx of digitized language available for analysis.

However, there is plenty of scope for us to get it wrong and for others to use our work in ways that harm individuals or groups.
What can we do?

Constant, pervasive surveillance leads to self-censorship. Yet, people also need public forums. How can we positively nudge governments and companies to select analyses that better protect individuals and their rights, regardless of legislation?

What research questions can we ask that make our work valuable, useful—even profitable!—while reducing the potential for harm to individuals or groups, or even promoting rights and societal good?

We must make choices, as individual researchers and as a field.
Work at People Pattern is in collaboration with Elias Ponvert, Joey Frazee, James Scott, Steve Blackmon, Abhishek Sinha and the rest of the team.

Thanks also to Bloomberg for sponsoring my travel!