The Human Manifold

On the Predictability of Human Online Behaviour and its Consequences

Thore Graepel

MSR Cambridge: Yoram Bachrach and Pushmeet Kohli
MSR Redmond: Kristin Lauter and Michael Nährig
Bing: Milad Shokouhi, Bin Bi
Cambridge University: Michal Kosinski and David Stillwell
Overview

- Predicting human attributes from online behaviour
  - Motivation and methodology
  - Results and interpretation

- Promises
  - Personalization of web services
  - Personalization of query auto-suggest

- Perils
  - Consequences for online privacy
  - Confidential Machine Learning

- Conclusions
Overview

• Predicting human attributes from online behaviour
  • Motivation and methodology
  • Results and interpretation

• Promises
  • Personalization of web services
  • Personalization of query auto-suggest

• Perils
  • Consequences for online privacy
  • Confidential Machine Learning

• Conclusions
The Digital Traces: Online Data

Social Media Landscape 2012

- Conversations
- Publishing
- Sharing
- Playing
- Localizing
- Buying
- Interacting
- Networking
- Connected devices

FredCavazza.net
High-Dimensional Observation Space
Mapping the Human Manifold

- **Scientific potential - Understanding:**
  - Better understand human behaviour
  - Understand commonality and individual differences among people
  - Obtain psychometric measurements at an unprecedented scale

- **Business potential - Predicting:**
  - Develop fine-grained and predictive psycho-demographic user profiles
  - Increase user satisfaction by deep personalization for products and services
  - Increase revenue by providing more engaging ads and recommendations

- **How: Find mapping to interpretable dimensions**
  - Personality, Intelligence, Happiness, etc.
### Big Five Personality traits

<table>
<thead>
<tr>
<th>Trait</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>• Appreciation of art, emotion, adventure, and variety of experience</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>• Self-discipline, act dutifully, and aim for achievement</td>
</tr>
<tr>
<td>Extraversion</td>
<td>• Energy, positive emotions, seek social stimulation</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>• Compassionate and cooperative rather than suspicious</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>• Experience unpleasant emotions easily, such as anger and anxiety</td>
</tr>
</tbody>
</table>
General Intelligence - IQ
Satisfaction with Life Scale (Happiness)

To what degree do you agree with the following statements?

- In most ways my life is close to ideal.
- The conditions of my life are excellent.
- I am satisfied with my life.
- So far I have gotten the important things I want in life.
- If I could live my life over, I would change almost nothing.
A Treasure Chest of Data: MyPersonality

- Facebook App since 2008
- Over 8 Million psychometric test results
  - Personality
  - Intelligence
  - Happiness
- Volunteered user profiles
  - Relationship status, age, gender
  - Facebook Likes
  - Friendship network

Data: [www.myPersonality.org](http://www.myPersonality.org)
David Stillwell, Michal Kosinski
Cambridge Psychometrics Centre
Your online reflection is built from your Big 5 personality test scores. It assesses key personality traits to get an overview of your character. It is currently the system most used by academic psychologists for personality research.

We have now compared you to more than 6.5 million previously gathered personality surveys, allowing the Mirror to predict the personality scores that you are likely to have.

Explore your personality scores and online behaviours through the links in the bottom left.

http://mirror.nationalmediamuseum.org.uk/
History of Psychometrics Work

- Pushmeet Kohli pioneers personality direction in project with intern Michal Kosinski, 2009
- Internal tech-report "Personality and Online Behaviour of the Crowd"
- Collaboration re-kindled, work resumed, 2011
- Bachrach, Kosinski, Graepel, Kohli, and Stillwell, "Personality and Patterns of Facebook Usage", in ACM Web Sciences 2012
- Kosinski, Stillwell, Kohli, Bachrach, and Graepel, "Personality and Website Choice", in ACM Web Sciences 2012
- Kosinski, Stillwell, and Graepel, "Private traits and attributes are predictable from digital records of human behaviour", in PNAS 2013
- Shokouhi, Bi, Kosinski, Graepel, "Inferring the demographics of search users", in WWW 2013

Fig. 6 Median of friends added per month by users characterized by different levels of Extroversion. Ribbon represents the interquartile range, or the middle 50 percentiles of the number of friends added per month.
Private traits and attributes are predictable from digital records of human behavior

Michail Kosinski1†, David Stillwell2, and Timm Groepel3

1Harvard Medical School, Boston, MA, 2Columbia University, New York, NY, 3University of California, San Francisco, San Francisco, CA

We show that many aspects of social records of behavior, Facebook and Twitter, can be predicted by traits and attributes that cannot be detected via traditional demographic questions such as age, sex, race, and education. We demonstrate that Facebook photos, Facebook likes, Twitter tweets, and demographic and psychological traits are strongly correlated. These findings provide new insights into the nature of social preferences and will pave the way for the development of novel interventions that can address the needs of individuals with high social risk.
The Magic of Machine Learning

[Diagram showing a user-likes matrix with 55,814 likes and 58,466 users.]

User - Like Matrix
(10M User-Like pairs)
High-Dimensional Observation Space
The Magic of Machine Learning

1. Users’ Facebook Likes

55,814 Likes

User 1 1 1 0
User 2 0 1 1
User 3 1 0 0
User n 1 1 0

User – Like Matrix
(10M User-Like pairs)

2. Singular Value Decomposition

100 Components

User 1 .5 .7 ... -.9
User 2 .3 -.4 ... -.2
User 3 -.6 .1 ... .7
User n 1.2 1 ... -.6

User – Components Matrix
Mapping the Manifold

100 out of 55,814 dimensions explain 28% of variance
The Magic of Machine Learning

1. Users’ Facebook Likes
   - 55,814 Likes
   - User – Like Matrix (10M User-Like pairs)

2. Singular Value Decomposition
   - 100 Components
   - User – Components Matrix

3. Prediction Model
   - Using Logistic or Linear Regression (with 10-fold cross validation)
   - \( e.g. \) \( age = \alpha + \beta_1 C_1 + \ldots + \beta_n C_{100} \)
   - Predicted variables
     - Facebook profile: age, gender, political and religious views, relationship status, proxy for sexual orientation, social network size and density
     - Profile picture: race
     - Survey/test results: Big5 Personality, Intelligence, Satisfaction with Life, substance use, parents together?
Making Predictions
Overview

• Predicting human attributes from online behaviour
  • Motivation and methodology
  • Results and interpretation

• Promises
  • Personalization of web services
  • Personalization of query auto-suggest

• Perils
  • Consequences for online privacy
  • Confidential Machine Learning

• Conclusions
Prediction Accuracy: Binary variables

- Gender: 0.93
- Lesbian: 0.78
- Gay: 0.88
- Democrat vs. Republican: 0.85
- Christianity vs. Judaism: 0.82
- Black vs. White: 0.95
- Uses drugs: 0.69
- Drinks Alcohol: 0.70
- Smokes Cigarettes: 0.73
- Parents together at 21: 0.60
- Single vs. In Relationship: 0.67
What is “Area under the Curve” (AUC)
How many Likes for good accuracy?

- About 50% of users in this sample had at least 100 Likes and about 20% had at least 250 Likes.
- Baseline (random guessing) is 50% for binary variables – Gender and Political View.
Prediction Accuracy: Numeric Variables

- **Age**: 0.75
- **Number of Facebook Friends**: 0.47
- **Density of Friendship network**: 0.52
- **Openness**: 0.43
- **Conscientiousness**: 0.29
- **Extraversion**: 0.40
- **Agreeableness**: 0.30
- **Neuroticism**: 0.30
- **Intelligence**: 0.39
- **Satisfaction with Life**: 0.17

**Correlation Coefficients**

- **Age**
- **Number of Facebook Friends**
- **Density of Friendship network**
- **Openness**
- **Conscientiousness**
- **Extraversion**
- **Agreeableness**
- **Neuroticism**
- **Intelligence**
- **Satisfaction with Life**
<table>
<thead>
<tr>
<th>IQ</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Godfather</td>
<td>Jason Aldean</td>
</tr>
<tr>
<td></td>
<td>Mozart</td>
<td>Tyler Perry</td>
</tr>
<tr>
<td></td>
<td>Thunderstorms</td>
<td>Sephora</td>
</tr>
<tr>
<td></td>
<td>The Colbert Report</td>
<td>Chiq</td>
</tr>
<tr>
<td></td>
<td>Morgan Freemans Voice</td>
<td>Bret Michaels</td>
</tr>
<tr>
<td></td>
<td>The Daily Show</td>
<td>Clark Griswold</td>
</tr>
<tr>
<td></td>
<td>Lord Of The Rings</td>
<td>Bebe</td>
</tr>
<tr>
<td></td>
<td>To Kill A Mockingbird</td>
<td>I Love Being A Mom</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>Harley Davidson</td>
</tr>
<tr>
<td></td>
<td>Curly Fries</td>
<td>Lady Antebellum</td>
</tr>
<tr>
<td>Satisfaction With Life</td>
<td>Satisfied</td>
<td>Dissatisfied</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td></td>
<td>Sarah Palin</td>
<td>Hawthorne Heights</td>
</tr>
<tr>
<td></td>
<td>Glenn Beck</td>
<td>Kickass</td>
</tr>
<tr>
<td></td>
<td>Proud To Be Christian</td>
<td>Atreyu (Metal Band)</td>
</tr>
<tr>
<td></td>
<td>Indiana Jones</td>
<td>Lamb Of God</td>
</tr>
<tr>
<td></td>
<td>Swimming</td>
<td>Gorillaz</td>
</tr>
<tr>
<td></td>
<td>Jesus Christ</td>
<td>Science</td>
</tr>
<tr>
<td></td>
<td>Bible</td>
<td>Quote Portal</td>
</tr>
<tr>
<td></td>
<td>Jesus</td>
<td>Stewie Griffin</td>
</tr>
<tr>
<td></td>
<td>Being Conservative</td>
<td>Killswitch Engage</td>
</tr>
<tr>
<td></td>
<td>Pride And Prejudice</td>
<td>Ipod</td>
</tr>
</tbody>
</table>
## Which Likes? Extraversion

<table>
<thead>
<tr>
<th>Extraversion</th>
<th>Outgoing &amp; Active</th>
<th>Shy &amp; Reserved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beerpong</td>
<td>RPGs</td>
<td></td>
</tr>
<tr>
<td>Michael Jordan</td>
<td>Fanfiction.Net</td>
<td></td>
</tr>
<tr>
<td>Dancing</td>
<td>Programming</td>
<td></td>
</tr>
<tr>
<td>Socializing</td>
<td>Anime</td>
<td></td>
</tr>
<tr>
<td>Chris Tucker</td>
<td>Manga</td>
<td></td>
</tr>
<tr>
<td>I Feel Better Tan</td>
<td>Video Games</td>
<td></td>
</tr>
<tr>
<td>Modeling</td>
<td>Role Playing Games</td>
<td></td>
</tr>
<tr>
<td>Cheerleading</td>
<td>Minecraft</td>
<td></td>
</tr>
<tr>
<td>Theatre</td>
<td>Voltaire</td>
<td></td>
</tr>
<tr>
<td>Flip Cup</td>
<td>Terry Pratchett</td>
<td></td>
</tr>
</tbody>
</table>
### Which Likes? Agreeableness

<table>
<thead>
<tr>
<th>Cooperative</th>
<th>Competitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compassion Intern</td>
<td>I Hate Everyone</td>
</tr>
<tr>
<td>Logan Utah</td>
<td>I Hate You</td>
</tr>
<tr>
<td>Jon Foreman</td>
<td>I Hate Police</td>
</tr>
<tr>
<td>Redeeming Love</td>
<td>Friedrich Nietzsche</td>
</tr>
<tr>
<td>Pornography Harms</td>
<td>Timmy South Park</td>
</tr>
<tr>
<td>The Book Of Mormon</td>
<td>Atheism / Satanism</td>
</tr>
<tr>
<td>Circles Of Prayer</td>
<td>Prada</td>
</tr>
<tr>
<td>Go To Church</td>
<td>Sun Tzu</td>
</tr>
<tr>
<td>Christianity</td>
<td>Julius Caesar</td>
</tr>
<tr>
<td>Marianne Williamson</td>
<td>Knives</td>
</tr>
</tbody>
</table>
Overview

- Predicting human attributes from online behaviour
  - Motivation and methodology
  - Results and interpretation

- Promises
  - Personalization of web services
  - Personalization of query auto-suggest

- Perils
  - Consequences for online privacy
  - Confidential Machine Learning

- Conclusions
How can we apply this to Bing users?

- Predictors are based on Facebook Likes
- Bing users are characterized by search history
- Train on common representation to transfer knowledge: Open Directory Categories
How good are transferred predictors?

- Age:
  - Facebook Likes: 77%
  - Bing Queries: 74%

- Gender:
  - Facebook Likes: 84%
  - Bing Queries: 80%
Political attitudes in the US

Gallup Survey

Predictions

2012 Election

Pearson Correlation: 0.72
Distribution of Christians in the US

Pearson correlation: 0.39
What can we do with this knowledge?

- Personalize our products and services based on inferred demographics & psychometrics
- Bing: Re-rank search results, query suggestions, query auto-suggests, ads
- Auto-Complete: Given query prefix, re-rank compatible query suggestions

<table>
<thead>
<tr>
<th>Features</th>
<th>MRR Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>+1.8 %</td>
</tr>
<tr>
<td>Gender</td>
<td>+2.6 %</td>
</tr>
<tr>
<td>Location</td>
<td>+3.1 %</td>
</tr>
<tr>
<td>Short history (session)</td>
<td>+0.9 %</td>
</tr>
<tr>
<td>Long history</td>
<td>+4.4 %</td>
</tr>
<tr>
<td>All features</td>
<td>+7.9 %</td>
</tr>
</tbody>
</table>

MRR: Mean Reciprocal Rank

Results: Milad Shokouhi, SIGIR 2013
Overview

• Predicting human attributes from online behaviour
  • Motivation and methodology
  • Results and interpretation

• Promises
  • Personalization of web services
  • Personalization of query auto-suggest

• Perils
  • Consequences for online privacy
  • Confidential Machine Learning

• Conclusions
Facebook threat to users’ privacy

By Andrew Levy

FACEBOOK users are at risk of unwittingly revealing personal details simply by ‘liking’ pages on the site dedicated to anything from celebrities to charities, researchers warn.

Sexuality, drug use, political views and religious beliefs can be accurately predicted by monitoring users’ activity on the social networking website, they said.

The team from Cambridge University focused their research on Facebook’s system of liking pages – the seemingly innocuous act of clicking a button illustrated with a thumbs up.

Worryingly, the researchers found that liking even apparently unrelated information still can be used to accurately predict personal details.

For example, the researchers found drug use is suggested by ‘liking’ milk shakes and swimming, while high IQs are indicated by showing a taste for curry fuses, and the Godfather movies.

The study was carried out by Cambridge’s Psychometrics Centre and based on the Facebook profiles of 50,000 people in the US.

Their ‘likes’ were fed into a computer algorithm which was used to predict a range of personality traits. Researchers predicted male sexuality with 88 per cent accuracy.

They also had an 89 per cent success rate with political leanings and 82 per cent with religion.

Dr Gus Hosein, of campaigners Privacy International, said: ‘It’s a nightmare scenario that Facebook is entirely responsible for setting up.

This information can be used to predict people’s likes and dislikes. People who are uncomfortable with this are well advised to stop using Facebook altogether.

Facebook declined to comment yesterday.”
Reactions

Sam Gosling, a psychologist at the University of Texas at Austin, calls it a "landmark study" because it illustrates "how things are no longer ephemeral." He has been studying Facebook behavior since 2006, and has seen this new study.

Dr Gus Hosein, of campaigners Privacy International, said: ‘It’s a nightmare scenario that Facebook is entirely responsible for setting up. This information can be used to pre-categorise people. ‘Banks could use it to decide who gets a loan. It also creates the perfect surveillance state for governments.’

less impressed. You already have more sensitive information online, Dr. Nicholas Christakis, director of the Human Nature Lab at Harvard, tells The Los Angeles Times. "I think this paper is alarmist. We can go from curly fries to pogroms in a couple steps."
Questions for User Privacy

- Are users aware to what degree we can infer their personal traits and attributes from their digital traces?
- If they were, would they make that data public?
- Do these predictions just summarize what users are signalling to their friends using Facebook Likes?
- Should companies/services be allowed to use this information for commercial purposes?
- Can technology help to mitigate these challenges?
Overview

• Predicting human attributes from online behaviour
  • Motivation and methodology
  • Results and interpretation

• Promises
  • Personalization of web services
  • Personalization of query auto-suggest

• Perils
  • Consequences for online privacy
  • Confidential Machine Learning

• Conclusions
3 Send encrypted values to Azure
2 Encrypt values with personal password
1 Enter medical data
4 Cloud runs prediction algorithm on encrypted data
5 Cloud returns encrypted prediction
6 Decrypt prediction with personal password

Machine Learning on Encrypted Data, Techfest demo with Kristin Lauter and Michael Nährig
Machine learning on encrypted data?

- **Yes:** Both training and prediction!
- **How:** Somewhat homomorphic encryption (SHE)
- **But:** Restricted class of machine learning algorithms:
  - Must be implemented as bounded-degree polynomials in the inputs
  - No division, no comparison
  - Just additions and very few multiplications
- **Prototype in computational algebra system Magma**
  - Linear means classifier
  - Fisher’s linear discriminant
### Proof-of-Concept: Breast Cancer Data

#### Linear Means Classifier: $q=2^{128}$, $t=2^{15}$, $\sigma = 16$, $n=4096$

<table>
<thead>
<tr>
<th># attributes</th>
<th># training vecs</th>
<th>Train (s)</th>
<th>Classify (s) per vec</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>60</td>
<td>1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>3.3</td>
<td>5.8</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2.6</td>
<td>2.0</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>6.2</td>
<td>6.0</td>
</tr>
</tbody>
</table>

#### Fisher’s Linear Discriminant: $q=2^{340}$, $t=2^{40}$, $\sigma = 8$, $n=8192$, 3-step gradient desc.

<table>
<thead>
<tr>
<th># attributes</th>
<th># training vecs</th>
<th>Train (s)</th>
<th>Classify (s) per vec</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60</td>
<td>3612</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>12212</td>
<td>20</td>
</tr>
</tbody>
</table>

Graepel, Lauter, and Nährig, "**ML Confidential: Machine Learning on Encrypted Data**" in *International Conference on Information Security and Cryptology – ICISC 2012*
Conclusions

- Mapping the manifold of human online behaviour
  - Science: Understanding human behaviour and large-scale psychometrics
  - User benefit: Deep personalization and adaptation to user preferences

- Balance need for privacy with quality of service

- Challenges and future work
  - Define a new predictive and interpretable framework for human behaviour
  - Which “Likes” do people get exposed to and how? Is there a filter bubble?
  - Discover causal relationships from observational data

- Collaboration with Bing and U Cambridge