# Digital records could expose intimate details and personality traits of millions

## 11. März 2013

***Research shows that intimate personal attributes can be predicted with high levels of accuracy from 'traces' left by seemingly innocuous digital behaviour, in this case Facebook Likes. Study raises important questions about personalised marketing and online privacy.***

Michal Kosinski

New research, published today in the journal PNAS, shows that surprisingly accurate estimates of Facebook users' race, age, IQ, sexuality, personality, substance use and political views can be inferred from automated analysis of only their Facebook Likes - information currently publicly available by default.

In the study, researchers describe Facebook Likes as a "generic class" of digital record - similar to web search queries and browsing histories - and suggest that such techniques could be used to extract sensitive information for almost anyone regularly online.

Researchers at Cambridge's Psychometrics Centre, in collaboration with Microsoft Research Cambridge, analysed a dataset of over 58,000 US Facebook users, who volunteered their Likes, demographic profiles and psychometric testing results through the myPersonality application.

Users opted in to provide data and gave consent to have profile information recorded for analysis. Facebook Likes were fed into algorithms and corroborated with information from profiles and personality tests.

Researchers created statistical models able to predict personal details using Facebook Likes alone. Models proved 88% accurate for determining male sexuality, 95% accurate distinguishing African-American from Caucasian American and 85% accurate differentiating Republican from Democrat. Christians and Muslims were correctly classified in 82% of cases, and good prediction accuracy was achieved for relationship status and substance abuse – between 65 and 73%.

But few users clicked Likes explicitly revealing these attributes. For example, less that 5% of gay users clicked obvious Likes such as Gay Marriage. Accurate predictions relied on 'inference' - aggregating huge amounts of less informative but more popular Likes such as music and TV shows to produce incisive personal profiles.

Even seemingly opaque personal details such as whether users' parents separated before the user reached the age of 21 were accurate to 60%, enough to make the information "worthwhile for advertisers", suggest the researchers.

While they highlight the potential for personalised marketing to improve online services using predictive models, the researchers also warn of the threats posed to users' privacy. They argue that many online consumers might feel such levels of digital exposure exceed acceptable limits - as corporations, governments, and even individuals could use predictive software to accurately infer highly sensitive information from Facebook Likes and other digital 'traces'.

The researchers also tested for personality traits including intelligence, emotional stability, openness and extraversion. While such latent traits are far more difficult to gauge, the accuracy of the analysis was striking. Study of the openness trait – the spectrum of those who dislike change to those who welcome it – revealed that observation of Likes alone is roughly as informative as using an individual's actual personality test score.

Some Likes had a strong but seemingly incongruous or random link with a personal attribute, such as Curly Fries with high IQ, or That Spider is More Scared Than U Are with non-smokers.

When taken as a whole, researchers believe that the varying estimations of personal attributes and personality traits gleaned from Facebook Like analysis alone can form surprisingly accurate personal portraits of potentially millions of users worldwide.

They say the results suggest a possible revolution in psychological assessment which – based on this research – could be carried out on an unprecedented scale without costly assessment centres and questionnaires.

"We believe that our results, while based on Facebook Likes, apply to a wider range of online behaviours." said Michal Kosinski, Operations Director at the Psychometric Centre, who conducted the research with his Cambridge colleague David Stillwell and Thore Graepel from Microsoft Research.

"Similar predictions could be made from all manner of digital data, with this kind of secondary 'inference' made with remarkable accuracy - statistically predicting sensitive information people might not want revealed. Given the variety of digital traces people leave behind, it's becoming increasingly difficult for individuals to control.

"I am a great fan and active user of new amazing technologies, including Facebook. I appreciate automated book recommendations, or Facebook selecting the most relevant stories for my newsfeed," said Kosinski. "However, I can imagine situations in which the same data and technology is used to predict political views or sexual orientation, posing threats to freedom or even life."

"Just the possibility of this happening could deter people from using digital technologies and diminish trust between individuals and institutions – hampering technological and economic progress. Users need to be provided with transparency and control over their information."

Thore Graepel from Microsoft Research said he hoped the research would contribute to the on-going discussions about user privacy: "Consumers rightly expect strong privacy protection to be built into the products and services they use and this research may well serve as a reminder for consumers to take a careful approach to sharing information online, utilising privacy controls and never sharing content with unfamiliar parties."

David Stillwell from Cambridge University added: "I have used Facebook since 2005, and I will continue to do so. But I might be more careful to use the privacy settings that Facebook provides."

For more information, please contact fred.lewsey@admin.cam.ac.uk

This work is licensed under a Creative Commons Licence. If you use this content on your site please link back to this page.

Digital records could expose intimate details and personality traits of millions