



Reverse analytics

Description

Can you infer a person's politics from their online footprint? Let's start with something less contentious: a person's nature preference. If you were raised in a school and home environment that encouraged you to spend time outdoors and enjoy nature pursuits, then it is highly likely that in later life you will respect the environment, enjoy being out in nature settings, and even enjoy looking at pictures of nature scenes.

Call those early life experiences A, and behaviours and attitude in later life B. In this case there's popular and research literature connecting A to B in a causal way. Though we are not 100% sure, in most cases people would accept that A implies B. A certain upbringing produces particular life habits. That's a logical *deduction*.

Now reverse the logic. You meet someone who talks a lot about the environment, likes outdoor pursuits and enjoys nature scenes. You might conclude that in early life they were raised in a nature loving environment and as a child spent a lot of time outdoors. But you can't be very sure of that. The causal links here become frayed. After all, someone could have had a nature epiphany in later life, decided to catch up on what they missed as a child, or yielded to the enthusiasms of a newly found nature loving social circle. So there's considerable uncertainty in this case in thinking that B implies A. But in so far as we attempt that inference, it comes under the category of logical *abduction*, a reverse kind of logical inference.

Certainty is low in such reasoning, and you would need more information to conclude something about a person's upbringing, especially if you don't want to ask them directly.

Social media data

To find out about an adult person's childhood, you could examine their Facebook feed (if they have one), and see how often they talk about or show pictures of outdoor pursuits, link to environmental causes, and even post pictures of nature settings. That would at least establish the B part of the logical equation. You may be able to find other evidence that hints at the kind of upbringing they had: anecdotes, advice, attitudes to children, etc. That would help establish the A part.

Now think of just one part of the B equation, e.g. preference for looking at, re-posting, or linking to pictures of nature. You could reasonably deduce (by looking, or using an algorithm) from a person's Facebook feed that they like nature pictures, more than pictures of cars, people, or buildings. Could you (or an algorithm) then infer (*abduct*) something about the person's upbringing? With the right detective skills you probably could.

Digital footprints

That's one of the intriguing and disturbing aspects of all that personal information many of us choose to put online. Our digital footprint reveals more about us than we state explicitly, and some of it can be gleaned from simple things like our choices of images.

What else could be *abducted* about us? It's not just whether we had a nature-loving upbringing, but our schooling, educational attainment, ethnic background, social circle, disposable income, purchasing habits, the kinds of holidays we take, alcohol consumption, personality profile and much else besides.

There is a [Cambridge University website](#) that claims to infer (abduct) your personality profile, age and other information just from an arbitrary segment of your writing, i.e. a blog, twitter feed or email – how you use language.

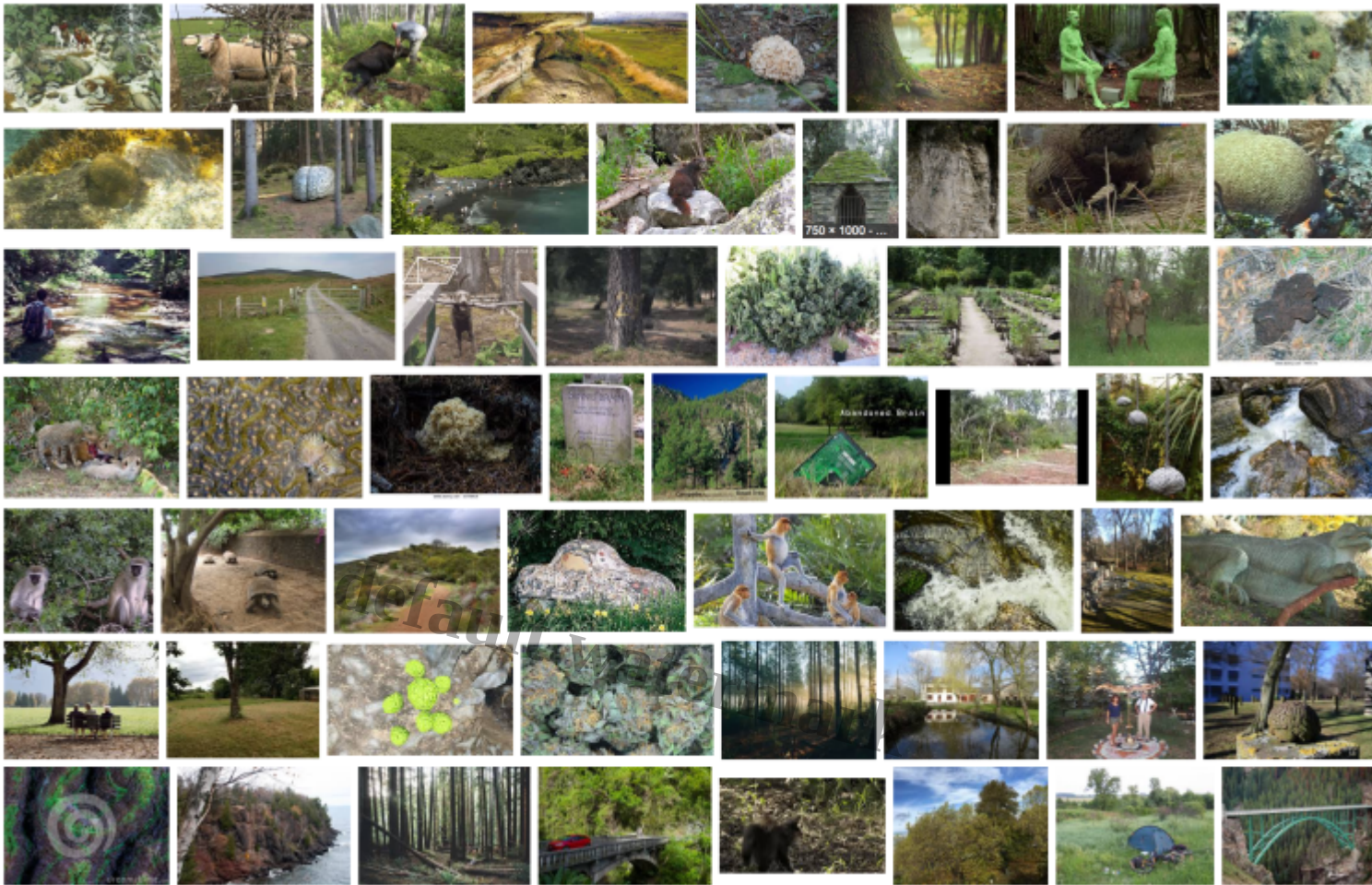
Approximate profiling

The methods deployed involve machine learning, or *induction* (the other mode of making inference) from a very large number of examples of writing by people who have completed personality tests, and whose profiles are already known. The system is far from accurate in assessing your personality or background, but it doesn't have to be. Accurate, individual personal profiling from your digital footprint is difficult (and threatening), but it's the aggregation of such assessments that delivers the impact.

Online targeted marketing makes use of this kind of abductive inference, and if automated, it only has to be approximately right about any individual's position within a demographic to impact large groups of people, and hence have an effect on profits.

Move that targeted marketing into political campaigning. If campaigners (or their consultants) target sufficient people with the message they want to hear, i.e. is tuned to voters' social and personality profiles, that could be sufficient to tip the vote in favour of one candidate over another. That's part of the opportunity, challenge and peril of big data analytics.

Also see posts: [Big data: a non-theory about everything](#), [Emotional targeting](#), and [Lego logics](#). See [Guardian](#) article for an update on [How social media filter bubbles and algorithms influence the election](#).



Bibliography

- Kitchin, Rob. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, (April-June)1-12.
- Louv, Richard. 2005. *Last Child in the Woods: Saving Our Children from Nature-Deficit Disorder*. London: Atlantic Books

Category

1. Nature

Tags

1. abduction
2. Big Data

Date Created

May 27, 2017

Author

rcoyne99