In the data: methods of interdisciplinary machine thought

Adrian Mackenzie

Sociology Department, Lancaster University

a.mackenzie@lancaster.ac.uk

abstract

How can we develop ways of thinking about and working with digital data that neither uncritically affirm belief in the power of data, nor reject beliefs in data as pure hype? This paper attempts do this on terrain that lie close to the centres of contemporary data practices: machine learning. The paper will illustrate some of the data practices and forms of knowledge associated with machine learning, an increasingly widely used way of programming computers to find patterns, associations, and correlations, and to make predictions on a large scale. It will discuss who is doing machine learning in industry, business and academia. Finally, it will discuss some of the problems and challenges in critically engaging with machine learning techniques for social sciences and humanities.

the problem

- much hinges on what we understand by `in' (James)
- find the function that generated the signal/data
  - Abbott
- cope with the data in width and length: real problem is width!

the data

11. The infrastructure --

a. Google compute i/o -- scaling up? why the genome?

12. In text
    a. e-discovery in US courts
    b. Text analysis program from UQ - Leximancer;
    c. Timothy Lin (Twitter) -- how it imprints;
13. In signal (images, sound)
    a. Lev Manovich -- cultural analytics;
    b. Google cats -- the difficulty;
    c. Eulerian motion

the people

1. Can social scientists do it?
    q. Gary King -- computational social science (King 2011)
    r. Savage -- descriptive assemblage
2. Can computer scientists do it?
    a. Andrew Ng -- the Stanford PhD's
    b. Programming Collective Intelligence
    c. Jaron Lanier - close to the server won't help
3. Can new hybrids do ti?
    a. Hilary from Bitly -- wonderful people
    b. Machine learning for hackers
    c. Cathy O'Neill/Heather Arthur/Rachel ??

the practices

9. OccupyData --
10. Programming
    a. R as magna carta
    b. hunch.com
11. The models
    9. Yihuir -- animation

References

King, Gary. 2011. ``Ensuring the Data-Rich Future of the Social Sciences." *Science* 331 (feb): 719--721. doi:10.1126/science.1197872. http://www.sciencemag.org/content/331/6018/719.abstract.