In the data: interdisciplinary modes of machine learning

Adrian Mackenzie

Sociology Department, Lancaster University

a.mackenzie@lancaster.ac.uk

Abstract

This paper explores ways of thinking about digital data that lie somewhere between blithe faith and critical dismissal. It focuses on the machine learning, an increasingly prevalent bundle of techniques and approaches that lies at the centre of contemporary data processing. Machine learning is used to program computers to find patterns, associations, and correlations, to classify events and make predictions on a large scale. As a set of techniques for classifying and predicting, machine learning lies close to centre of calculation in social network media, finance markets, robotics, and contemporary sciences such as genomics and epidemiology. This paper will discuss who is doing machine learning, who could do machine learning, and how they might do it differently.

The problem

- finding the function that generated the data
- pattern finding vs models
  - Abbott - inaccurate opposition between linear and pattern

-

The data

- google compute

  - google io -- why genome? wide and mixed; scaling up
- text data
  - James Lin, Twitter
  - e-discovery
  - leximancer
- signals -- image and sound
  - google cat
  - driverless car - Thrum

The people

- the computer scientists
  - Ng: stanford Phds
  - programming collective intelligence
- the wonderful people
  - Hilary
  - Cath O'Neill & Rachel whats her name
  - Heather Arthur
- the social scientists
  - Gary King
  - machine learning for hackers
  - manovich -- cultural analytics
  - Savage - descriptive assemblage

What to do

- Jaron Lanier
- occupydata
- animation
- text/coding/reproducibility

References