

Module 3: big data, machine learning and the social sciences

Sunday, June 11, 2023

20:09

- Fairness, accountability and transparency
- Computational social sciences - to study social process
- data:
 - Big data - amassing huge amounts of statistical information on social and economic trends and human behavior (social data sets)
 - Its not about size its about granularity - granularity of individual people and activities
 - Brings up issues of privacy, bias, etc.
 - new issues fall outside expertise of those aggravating. And analyzing data
 - Need to stop thinking of big data as homogenous - data sets nested
- questions:
 - Fine grained patterns may not be readily visible using existing computational techniques
 - data-first and method first approaches can amplify issues
 - "convenience" data sets typically reflect only a particular privileged segment of society
 - Need to think about best instruments data aggregation and curation mechanisms - for fine graded patterns
 - Algorithmic accountability reporting work in journalism
 - instead of engaging in data-first or model-first research and development, prioritize questions - standard data set may not be appropriate
- models:
 - Many are productive models
 - Computer scientists - find needle haystack (night webpageto display)
 - Social scientists are more commonly interested in characterizing the haystack
 - Three categories of modeling:
 - prediction:
 - Observed data to make predictions about missing information

(compsci)

- Logistic regression, conditional random fields, naive bayes, gaussian process and support vector machines

- Explanation:

- Finding plausible or probable explanations for observed data (social scientists)
- Compared with established social theories or previous findings.
- Characterizing the haystack

- Explorations:

- Uncovering patterns in observed data
- What does the data tell me I don't already know
- Latent dirichlet allocation, factor analysis and stochastic block models
- Rely on existing knowledge - previous explanatory or exploratory analyses (bias intro)
- Models used to perform previous analysis will necessarily influence the resultant findings and representations we choose to use in predictive models

- Dangers of unnoticed errors
- Big difference between 95% accurate because of noise and one that only nails classifying for one subgroup - need careful error analysis
- Preventing or mitigating error - explicitly represent and maintain uncertainty - less confident predictions will be about minority groups
- Even correct predictions can be predicated on weak data sets
- report, don't ignore uncertainty
- Model validation and error analysis are more tricky for exploratory models - most don't have a single answer like predictive models
 - Multifaceted approach to model validation by identifying different qualitative and quantitative tasks each intended to validate a particular aspect

- findings: in

- Cognitive shortcuts that unconsciously influence
- Confirmation bias: confirms areas conceived hypothesis
- search for confirmatory and contradictory evidence for hypothesized findings
- Increase public understanding of science

