notes

Methods for for Studying Algorithmic Systems

Motivating Examples

{Ask to skip. Likely enough in previous sessions.}

Discrimination

- [SaW] "jew" vs "jewish people" disclaimer after complaint at Anti Defamtion League. "Couldn't do more as it's computer-generated"…later in article: fixing results should be doable. filtering neo-nazi material in France and Germany
- [SaW] "black woman" → "black girls porn" (e.g. via duckduckgo due to porn filter that can be statically disabled. via google → 1st: "Angry Black Woman", 3rd: "20 Popular White Celebrities Who Have Black Spouses"). instead of not *culturally situated / articles written by black women*
- [SaW] "women's magazines" (via google) → no feminist literature to be found (would give women confidence). entirely big magazines like Cosmopolitan & Women's Day (all by Hearst Corp). Not even a higher weighting for *variety* of perspectives.
- [SaW] "women atheletes" (via google) → "top 25 sexiest female athletes"
- [SaW] own search: "unprofessional hair" (via google images) → black women
- [SaW] own search: "woman" (via google images → white women
- [SaW] "Black Girls Rock!" campaign managed to reach first page. had a lot of budget.
- [SaW] If we don't fix these: "identity-based search results could be nothing more than old bigotry packaged in new media."
- [Rge] result for searches with "black" names in online background-check service Instant Checkmate contain "arrest" more often

- [AAg] price discrimination [Rge] e.g. by location, as done by e-commerce plattform staples.com
- [AAg] economic opportunity (ressources and information)

- [AAg] fair housing chances in housing search

Antitrust

[SaW]

- money & lot of domains/content for crosslinking → Search Engine Optimisation
- bought keywords
- ad-space

[SaW] google seems to prioritize own subsidiaries (youtube over others, google maps over mapquest and osm, google images over photo

[AAg] issues:

- commercial rigging
  - e.g. American Airline's SABRE (regulated after antitrust), Google (Health) over e.g. WebMD
  - while projecting notion of objectivity
  - appear like good results while being anti-competitive
- payola

Misc

[SaW] high-ranking of pages leads to trust

[Rge] search engine auto-completion censors (value-loaded choices)

[Opc] spam false-positives get censored (e.g. legit email send from false country of origin ← studied but found no evidence. containing wrong words like "guarantee", "plea", "dollar", "visit")

[Rge] False positives/negatives: Yelp filters vs bought reviews but also censors some legit ones in turn

[AAg] practices might not always be illegal but might have negative societal/economic consequences

[Opc] classification systems valorize some points of view and silence others. can break lives (e.g. apartheit, credit evaluations, feministic magazine (see "search-engines about women" above), …)

Challenges

Constantly changing

[HtS] Algorithms are adapted over time (ontogenetic and perfomative)

Algorithms are out of control

- [HtS] unexpected consequences
  - e.g. google results from "Motivations"-section (assuming accident over intentional malice/racism)
- [HtS] unexpected usages
  - [WwM] e.g. uber-drivers turning app on/off to game assignment and reward algorithms

Intentional Opacity / Black Boxing

[HtS][Opc][AAg] trade secret / intellectual property

[Opc][HtS][AAg] "gaming the system" cat-and-mouse

- vs search engine optimisation and google bombing)
- e.g. reddit-core is closed (rest is open)

[Opc] vs scrutiny and legislation

[Opc] need for trusted auditors and supporting legislation

Opacity due to technical illiteracy

[Opc] code both for humans and machines / code quality

[Opc] need schooling (e.g. algorithmic journalists) and cooperations (technical, legal, journalistic,…)

Scale Opacity

- [Opc] large datasets with many dimensions
  - statistics from 2014 → google and ebay ~100PB per day(!)
  - Eric Schmidt: "Every 2 Days We Create As Much Information As We Did Up To 2003" quote from techcrunch
- [HtS] large teams
  - statistics from 2015 → google: ~30,000 engineers
- [Opc] huge, heterogenous codebases
  - [HtS] pre-existing protocols and frameworks/libraries
  - [HtS] networks of algorithms (services?)
    - google codebase has roughly 4 * 10ˆ12 lines of code (see http://www.informationisbeautiful.net/visualizations/million-lines-of-code/)

Implementation Opacity - Curse of Dimensionality

[Opc] curse of dimensionality / high-dimensionality in itself and due to non-human-readable encoding

- e.g. principal component analysis
- e.g. kernel trick in SVMs

[Opc] prefer feature extraction over pca/kernel trick

Implementation Opacity - Non-human-readable encoding

[Opc] decentralized / non-human-readable encoding / algorithm doesn't think along human lines

- e.g. bag of words @ spam-filtering: high-level scam-patterns vs bag of words that can easily be fooled via synonyms and is obscure in it's decisions)
- e.g. neuronal networks

[Opc] need better interpretability (necessary for accountability but also technical imrovements)

[Opc] or avoid machine learning in certain critical dimensions

Methods

{Ask for for familiarity → level of detail}

[Rge] asks for algorithmic journalism

[Opc] argues for study of classes of algorithms instead of specific implementations+contexts to surface broader patterns/risks

Studying the people

- [HtS][Rge] interviews and ethnographies of **coding teams**
  - → motivations, driving values and methods
  - no obligation for truth!
- [HtS] studying the effects of algorithms on **users** and **contexts**
  - [HtS] via user experiments, interviews, ethnographies
  - [HtS] contexts: study legal, economic, institutional, technological, bureaucratic, political (etc)

Audit Studies in Social Sciences

{skip?}

[AAg] used for studying discrimination in social sciences and court cases

[AAg] two equivalent appliciations, different racial/gender/… identies (e.g. via name)

[AAg] are field studies:

- "research design involving the random assignment to groups in a controlled setting in order to isolate causation"
- more uncontrolled variables → causal inference more difficult
- results better generaziable to real-world-settings

[AAg] types:

- correspondence
- in-person (using actors)

[AAg] ethical challenges:

- no preliminary, informed consent – while binding their ressources (e.g. time)
- predisposed to prove guilt of studied person/group
- nevertheless became an accepted practice due to the valuable evidence gained

Studying the Algorithm (Algorithm Audits)

[AAg] difference: focused on one/few plattform(s) because:

- they have huge market-shares
- often: comparing apples and oranges

ethical challenges (see "Audit Studies in Social Sciences")

Manual Black Box Analysis

[HtS][Rge] reverse-eengineering / black-box analysis via varied input

limit: not much data → causation might not be clear

Reflexively producing code

= "implement it yourself"

- [HtS] auto-ethnographic ~ journaling
- [HtS] extremely vulnerable to biases! → use for exploration in conjunction

Reading the Source (Code Audit)

- [HtS] examining pseudo-code, source code and documentation
  - [HtS] reading code and docu (deconstruction)
  - [HtS] mapping changes (genealogical mapping)
  - [HtS] different languages/frameworks/… (comparative analysis)
  - [HtS] prob: messy code / huge codebases

[AAg] open source (or shared with independent evaluators)

[AAg] limits:

- see "black-box" above (IP, gaming the system)
- huge code-bases
- discrimination not localized but all over place (`if ($race = NOT_CAUCASIAN) then { illegal_discrimination() };`
- dynamic pages (might no be generated twice)

Survey Users ("Noninvasive" User Audit)

[AAg] **survey** users

[AAg] ask to install **tracking** software

[AAg] limits:

- need very diverse set of participants to detect bias in diff
- expensive
- self-reporting vs behaviour
- demand bias (e.g. answering "what you want to hear")
- not for sensitive areas (e.g. health and finance)

Scraping

scrape everything and **explore** (e.g. via clostering)

[AAg] e.g. netflix scrape for micro-genres

[AAg] limits:

- not possible for personalised pages (e.g. google results)
- login-walls
- way of accessing differs from regular usage
- protections vs bots and scraping (obfuscation, captchas,…)
- no randomization/manipulation to be used as independ variable (preliminary hypothesis)
  - causation not clear
  - problems with biases
- legal problems (w/o consent by provider)
  - US Computer Fraud and Abuse Act

- Terms of service

Sock Puppet Audit

independent variables: by varying input (e.g. location, gender identity, …)

[AAg] script user-profiles and interactions → great deal of control

[AAg] good for non-public features

[AAg] limits and challenges:

- legal problems
- how much likeness to humans needed?
- might influence algorithms with small user-bases, skewing results

Crowd-sourced / collaborative audit

independent variable: different users (e.g. location, gender identity, …), different tasks, etc

[AAg] Hire many human testers (e.g. via Amazon's Mechanical Turk). Fixes legal problems and problems with human-likeness of sock-puppet audit. Has side-effect of creating a problem-aware community.

[AAg] e.g. BiddingForTravel - "average" price

[AAg] limits and challenges:

- needs concise tasks e.g. "perform task, save result-html and send it to us" (not much possibility for feed-back)
- expensive

Misc

{skip}

[Emp]…importance of empirical studies when determining perfomance / algorithmic complexity.

[Emp]…empirical approach requires rigorous experimental design and empirically-based explanatory theories.

Discussion Points

problems with methods in papers so far(?)

Self-regulation: "bias testers" similar to penetration testers? Benefit to companies?

From “Auditing Algorithms”:

- Avoiding influencing (smaller) platforms? – “How difficult is it to audit a platform by injecting data without **perturbing the platform**?”
- Minimum amount of data to detect bias? – “**What is the minimum amount of data** that would be required to detect a significant bias in an important algorithm?”
- What proofs/evidence for the public? – “What **proofs** or certifications of algorithmic behavior could be brought to bear on public interest problems of discrimination?”
- What behaviour do we want? – “Accountability by auditing: how do we as a society **want** these **algorithms to behave**?”

References

- **[SaW]** … [4 pages] Noble, Safiya. 2012. “*Missed Connections: What Search Engines Say about Women.* Bitch magazine , 12(4): 3741. https://safiyaunoble.files.wordpress.com/2012/03/54_search_engines.pdf
- **[Rge]** … [blog article] Diakopolous, Nick. 2013. *“Rage against the Algorithms”* The Atlantic, October 3. http://www.theatlantic.com/technology/archive/2013/10/rage against the algorithms/280255/
- **[HtS]** … [blog article] *How to Study Algorithms: Challenges and Methods* https://algocracy.wordpress.com/2016/03/14/how to study algorithms challenges and methods/
- **[AAg]** … [18 pages] Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. *“Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms.”* Data and Discrimination: Converting Critical Concerns into Productive Inquiry , 64th Annual Meeting of the International Communication Association. May 22, 2014, Seattle, WA, USA.
- **[Opc]** … [18 pages] Burrell, Jenna. 2015. *“How the Machine ‘Thinks:’ Understanding Opacity in Machine Learning Algorithms.”* http://bds.sagepub.com/content/3/1/2053951715622512
- **[Emp]** … [23 pages] Hooker, J.N. 1994. “Needed: An Empirical Science of Algorithms.” Operations Research 42(2): 201212. http://www.akira.ruc.dk/~keld/teaching/algoritmedesign_f08/Artikler/01/Hooker93.pdf
- **[WwM]** … [10 pages from “Algorithmic Culture”-section] Lee, Min Kyung, Kusbit, Daniel, Metsky, Evan and Dabbish, Laura. 2015. “*Working with Machines*: The Impact of Algorithmic and Data-Driven Management on Human Workers” CHI ’15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems: 1603-1612 h ttp://dl.acm.org/citation.cfm?id=2702548