Section Eight

Class COGS 9

Teaching Assistant: Matthew Feigelis

DOI:10.1145/2844110

Article development led by acmqueue
queue.acm.org

A view from computational journalism.

BY NICHOLAS DIAKOPOULOS

Accountability in Algorithmic Decision Making

EVERY FISCAL QUARTER, automated writing algorithms churn out thousands of corporate earnings articles for the Associated Press based on little more than structured data. Companies such as Automated Insights, which produces the articles for the AP, and Narrative Science can now write straight news articles in almost any domain that has clean and well-structured data: finance, sure, but also sports, weather, and education, among others. The articles are not cardboard either; they have variability, tone, and style, and in some cases readers even have difficulty distinguishing the machine-produced articles from human-written ones.[4]

It is difficult to argue with the scale, speed, and labor-saving cost advantage that such systems afford. But the trade-off for media organizations appears to be nuance and accuracy. A quick search on Google for "'generated by Automated Insights' correction'" yields results for thousands of articles that were automatically written, published, and then had to have corrections issued. The errors range from relatively innocuous ones about where a company is based, to more substantial wrong word choices—*missing* instead of *beating* earnings expectations, for example. Were any of these market-moving errors? Was the root cause bad data, a faulty inference, or sloppy engineering? What is the right way to post corrections?

Algorithmic curation of content is also behind some of the most important and influential news-dissemination platforms that virtually all of us use. A recent Pew study found Facebook is a source of news about government and politics for 61% of millennials,[18] yet a majority of the public is not aware the Facebook newsfeed is algorithmically curated.[11] This becomes a lot more problematic when you consider Facebook can affect voter turnout in elections based merely on the amount of hard news promoted in an individual's news feed.[24] This bit of information, together with recent research showing biased search results can shift the voting preferences of undecided voters,[10] points to the need to start asking questions about the degree to which such curation and ranking systems can affect democratic processes.

These are just a few examples of algorithms influencing our media and information exposure. But the impact of automated decision making is being felt throughout virtually all strands of industry and government, whether it be fraud-detection systems for municipalities managing limited resources, a formula that grades and ranks teacher performance, or the many ways in which dynamic product pricing is done by Amazon, Airbnb, or Uber.[8] It is time to think seriously about how the algorithmically informed decisions now driving large swaths of society should be accountable to the public. In the face of important or expensive errors, discrimination, unfair denials of public services, or censorship, when and how should algorithms be reined in?

Computer science and engineering

Background

Goal of the paper: Identify some issues with machine learning and suggest improvements to accountability and transparency mechanisms in private and public sectors

- Algorithm's are used in government and industry. E.g., Algorithmic curation of content in twitter, facebook. Dynamic product pricing is done by Uber.
- How can we improve the the positive effects of these algorithms on our lives?
- Author presents mechanisms for accountability and transparency that may help

Five Elements of Transparency

Human element

- What is the goal of the algorithm?
- Who developed the algorithm?
- Who at the institution has control over the algorithm?

Data

- What features or variables are used in the algorithm?
- How were the variables defined, collected, processed?
- Is the data public or private?

Five Elements of Transparency

Model

- What are the assumptions (statistical or otherwise) behind the model?
- What models were tested, and why was the particular one chosen?

Inferencing

- What is the accuracy of the model?
- How many false positives versus false negatives are there?
- What kinds of steps are taken to remediate errors in the event of one?

Algorithmic presence

- What algorithms are being employed on the user? For example, A/B testing, filtering, personalized recommendation

Our Opinion

- What, of these previous questions, should the institution be required to disclose?

- What are the concerns of disclosing some information?

- What are the benefits?

Discussion Question

Discuss amongst groups of 2-4 people the following. Take 5 minutes. Then we'll reconvene

- Who's responsible for the algorithms effects? The developers? The managers? The CEO's?
- Should government and industrial leaders know the ins and outs as well as the programmer / creator does?

Attendance

- Enter your number and today's word into the attendance form

- Todays word: Boba
- Form: https://forms.gle/tx9GcpANHEMwj8Jv7

- https://tinyurl.com/cog9-spring-23

Office Hours

- Final Project Part 2 due Fri, Jun 09