CAPS Case - Lecture 4
Auditing Algorithms

Leo Klenner, Henry Fung, Cory Combs

Last updated: 17/9/2019

Goal

This is the case we will discuss in the fourth lecture. The goal of this case is to familiarize you with the challenges of auditing an algorithm.

One of the core challenges of auditing an algorithm is that you may not have access to the algorithm's code, or "inside", because it's guarded by the company, organization or state that developed the algorithm. Instead, you have to work with the algorithm's observed behavior, or "outside"; to build an understanding of how the algorithm works, you have to observe how the algorithm responds to your input, how this compares to other algorithms, etc. This is a complex task because your methods of inquiry are inherently fuzzy and thus you have to make a judgement about the algorithm based on imperfect information. But it's an important task, because the workings of algorithms are often opaque or hidden and, in the interest of users or other stakeholders, we may want to achieve the greatest transparency possible.

The case is based on a recent investigation of the Wall Street Journal into Google's search algorithms, ["How Google Interferes with its Search Algorithms and Changes Your Results"](). A brief summary of this investigation is given in the next section.

In the case, you have to take two sides:

- First, you have to review the methodology that the Journal used to audit the "outside" of Google's algorithm and make recommendations for improvements of the methodology.
- Second, you have to review the methodology that Google uses to adjust its search algorithms and make recommendations for improvements of the methodology.

Side note: The case also introduces two other themes of lecture 4. The first theme is a machine learning technique called *reinforcement learning* as the suggestion and ranking of search results can be framed as the problem of personalized recommendations, which is commonly solved through a reinforcement learning framework called multi-armed bandits. The second theme is *how humans and algorithms can interact* to arrive at more controlled decision-making as the Journal's investigation centers around Google's use of humans to review and adapt the performance of its search algorithms.

Background

Premise

Here is the premise of the Journal's investigation ["How Google Interferes with its Search Algorithms and Changes Your Results"]():

> The company states in a Google blog, "We do not use human curation to collect or arrange the results on a page." It says it [can't divulge details about how the algorithms work](#) because the company is involved in a long-running and high-stakes battle with those who want to profit by gaming the system.
>
> But that message often clashes with what happens behind the scenes. Over time, Google has increasingly re-engineered and interfered with search results to a far greater degree than the company and its executives have acknowledged, a Wall Street Journal investigation has found.

The goal of the journal's investigation is to show that humans interfere with Google's search results (ie. certain sensitive search results that you might expect to find, can't be found).

Findings

Here are some of the findings of the Journal's investigation:

> • Despite publicly denying doing so, Google keeps blacklists to remove certain sites or prevent others from surfacing in certain types of results. These moves are separate from those that block sites as required by U.S. or foreign law, such as those featuring child abuse or with copyright infringement, and from changes designed to demote spam sites, which attempt to game the system to appear higher in results.
>
> • In auto-complete, the feature that predicts search terms as the user types a query, Google's engineers have created algorithms and blacklists to weed out more-incendiary suggestions for controversial subjects, such as abortion or immigration, in effect filtering out inflammatory results on high-profile topics.
>
> • To evaluate its search results, Google employs thousands of low-paid contractors whose purpose the company says is to assess the quality of the algorithms' rankings. Even so, contractors said Google gave feedback to these workers to convey what it considered to be the correct ranking of results, and they revised their assessments accordingly, according to contractors interviewed by the Journal. The contractors' collective evaluations are then used to adjust algorithms.

The findings concern use of keyword blacklists to interfere with search results and auto-complete suggestions as well as the use of human contractors to evaluate and adjust Google's search results.

Tasks

Review of the Journal's Methodology

First, you are tasked with reviewing the methodology that the Journal used to audit the "outside" of Google's algorithm, which is stated below.

Answer the following questions:

- Drawing on your knowledge of machine learning algorithms, do you think the Journal's method of finding evidence of human interference in Google's search algorithms (blacklisting) is conclusive?
- How would you improve the Journal's methodology to obtain stronger evidence of human interference, given that you cannot access the algorithms' "inside"?

The Journal's methodology

Summary

In brief, the Wall Street Journal claims to find evidence of human interference in Google's algorithms through the following comparison of three search engines, Google, Bing, and DuckGoGo:

- given the same input keyword such as "Joe Biden is", what are the differences in auto-suggestions, such as "running for President" between the search engines?

Details on the methodology

> As part of its examination, the Journal tested Google's search results over several weeks this summer and compared them with results from two competing search engines, Microsoft Corp. 's Bing and DuckDuckGo, a privacy-focused company that builds its results from syndicated feeds from other companies, including Verizon Communications Inc. 's Yahoo search engine.
>
> The testing showed wide discrepancies in how Google handled auto-complete queries and some of what Google calls organic search results—the list of websites that Google says are algorithmically sorted by relevance in response to a user's query.

> We deployed code [...] that would mimic a human typing a phrase into a query box, such as "Joe Biden is." The resulting auto-complete suggestions from each search engine were captured by recording the HTML, the code that represents the content of a webpage.

Note, that the Journal released a detailed guide to its methodology, in which it states that:

- searches were performed by algorithms mimicking a human typing
- before each search, a new IP address was generated
- all the computers that ran the algorithms were based in the same location, Virginia
- the testing took place over different dates to look at various terms that were relevant in the news
- on each of these dates, searches for a selected term were performed every two hours

Example of comparison between search engines

Here is an example of the comparison between Google and Bing auto-completes for searching "Joe Biden is":

Joe Biden is ▼

GOOGLE		BING	SHOW DUCKDUCKGO
done	100%	donald trump	100%
how old	100%	a sen	78%
from	99%	he done	78%
running for president	79%	a reclamation project	71%
he democrat	78%	presidential	64%
he running for president	76%	a grouper	58%
toast	71%	going to cure cancer	58%
a democrat	70%	issues	58%

Figure 1: Comparison between Google and Bing auto-completes for "Joe Biden is"

Review of Google's Methodology

Second, you are tasked with reviewing the methodology that Google uses, according to the Wall Street Journal, to interfere with its search results. The method is stated below.

Answer the following questions:

- Drawing on your knowledge of machine learning algorithms, do you think Google's method of interfering with its search results is sound?
- How would you improve Google's method? To answer this question, consider how you would specify Google's goal on this matter.

Details on the methodology

Note, that this is the same text excerpt included in the background section.

> • To evaluate its search results, Google employs thousands of low-paid contractors whose purpose the company says is to assess the quality of the algorithms' rankings. Even so, contractors said Google gave feedback to these workers to convey what it considered to be the correct ranking of results, and they revised their assessments accordingly, according to contractors interviewed by the Journal. The contractors' collective evaluations are then used to adjust algorithms.