CS510: Explorations in Data Science Jeff Thomson
July 24, 2020 Aaron Hudson

Bias in Machine Learning Algorithms

Midpoint Report

Midpoint Meeting with Kristin: July 23, 2020 at 1:30pm

Project Objective:

With the ever-increasing utilization of AI in our society to make decisions that directly impact humans, monitoring the fairness of these AI systems is of the utmost importance. AI is often viewed as a method to remove human-bias from decision-making, employing a more clinical approach to handling data than humans are capable of. Interestingly enough, AI can oftentimes enact and reinforce human biases in covert ways. We plan to approach this topic from an educational perspective of making people aware of what biases look like in regards to AI and how they manifest by exploring some real-world examples. Then, we will examine methods that can be used to identify potential issues that may lead to bias or to determine if an algorithm is acting in a biased way, and will tie them back into the examples from before. Finally, we may explore some techniques utilized to counteract bias within our examples.

Project Approach:

Since we are approaching this topic from two major facets (examples of bias, and options for how to identify them) our research will consist of finding necessary information to expand upon these.

First, we plan to find specific examples of bias in AI and explore how they occurred. This will include locating articles referencing a few instances of AI that exhibited biased behavior and analyzing what led to them learning the bias. Some specific examples we will be focusing on are hiring algorithms, sentiment analysis, facial recognition, chat bots, and algorithms for determining crime rates.

After exploring these examples, we will look at ways the bias within them was determined, and possibly ways it could have been predicted. Furthermore, in cases where it is applicable, we will also talk about ways to compensate for the bias present. A very nice source for methods of counteracting bias is the Fairness, Accountability, and Transparency in Machine Learning website (https://www.fatml.org/). It contains a plethora of papers that identify methods used to circumvent bias in a variety of scenarios.

Our initial plan for researching is to work independently, as to not bias one another in how we approach the topic and to generate the most information possible. We will then share the sources found to review them together and finalize our choices for use in our paper. During the selection process, we will determine if there are any gaps in our knowledge base, and work together to find any remaining sources necessary.

Team Structure:

Our game plan is to search for sources independently at first and then to come together to share our findings. Once our sources have been selected, we will break down writing the paper into each of us tackling a set of the examples we will be evaluating. This will consist of us each writing the initial evaluation of the real-world example and writing the section about how the bias was identified and could be compensated for (if applicable). Then, once our individual sections have been finished, we will combine them together and ensure they flow coherently, and will finish by writing the introduction and conclusion.

Milestones:

1. Research and sources required for paper located
2. Outline of paper finalized and agreed upon
3. Midpoint Report completed
4. Rough draft of paper written
5. Paper finalized and completed
6. Presentation for class developed from paper

Challenges Encountered:

While bias is a very hot topic in machine learning, counteracting it is still an area that is not as heavily fleshed out. Finding research on it has been difficult that relates to real-world examples, as in many real-world cases the companies involved have kept their algorithms private. They have mostly released information on the overall findings, while omitting many of the details. Due to this, we have shifted the focus from compensating for and correcting bias to concentrate more on ways to identify it. This will be approached from the standpoint of predicting bias before the algorithm has even been run, to also how to look at the results it provides and identify bias within them. Furthermore, we plan to include references to additional readings if people are interested in learning about how to mitigate bias in ways that don't relate specifically to our examples chosen. We feel that being able to identify bias is an important skillset and one that lends itself well to an educational paper.

Progress Update:

At this point we have located a majority of our sources and have a rough outline of our paper finalized. We are still working on the finalized outline, as we want to ensure we have enough material for the length requirement. We have also made a few shifts on the focus of our paper and need to locate some sources for those changes. Specifically, after talking with Kristin at our Midpoint Meeting, we are going to try to focus more on specific examples and relating each of our concepts to those real-world examples to make the paper more engaging and memorable for our audience. Additionally, we are going to shift from solely focusing on rectifying bias to also include identifying it, both preemptively and identifying it based on results from algorithms. This is to account for the difficulty of finding techniques for how to rectify bias.

We also made some modifications to our project plan (as outlined above) to account for changes we have made to the overall project objective and approach.

We also changed our team structure to allow each of us to focus on specific examples of bias, rather than one person focusing on examples and the other focusing on how to identify/rectify them. We felt this was a more natural breakdown of topics, as we could each deep-dive into our assigned examples to allow a more thorough understanding and analysis of them.

We kept our milestones the same as our project plan, and while we wanted to have our sources and outline finalized by this point, recent shifts in the direction of our paper require we spend a bit more time in those areas. That said, we are very close to being done with those areas and will be ready to work on our paper after that. Furthermore, much of our R/P paper will be utilized for our final paper, as it serves as an introduction to our topic.