Thesis Project Portfolio

Quantifying Uncertainty of V-Information for Data Valuation

(Technical Report)

Sources of Bias in Machine Learning Models and Methods to Mitigate Them

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Srinivasa Josyula

Spring, 2023

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Introduction

In recent years, the field of machine learning has grown vastly. Moreover, many subfields of machine learning such as natural language processing and computer vision have risen. These technologies allow computers to complete human-like tasks such as facial recognition and language translation. However, these technologies are very complicated and they rely heavily on past data to learn and solve future tasks. My STS research focuses on the effects of biased datasets on creating biased models. In this research, I analyze various cases of biased models to determine short and long term solutions to improve equality. My technical research focuses on improving a current metric that computes the amount of information in a dataset. This metric can be used to test the effectiveness of different datasets on one model. My STS and technical project are loosely related because this metric could potentially be used to determine the source of bias in large datasets.

Project Summaries

My STS research was focused on determining the cause of biased machine learning models and potential solutions to them. I analyzed numerous cases of bias found in facial recognition, voice assistant and medical detection technology. This made it clear that the main source of bias came from biased datasets. In most cases, machine learning models were not trained on equitable datasets so they provided unequal results for different groups. Any potential solution for this would involve a process of creating more equitable datasets and re-training the models. This process would require collaboration from software developers and experts from the specific field to ensure a long term equitable solution. In my research, I pinpointed specific steps that must be taken in each of the above mentioned scenarios.

My technical research involves improving a current metric that measures the amount of usable information in a dataset. This metric known as v-information allows researchers to compare the performance of different datasets on a model. A closely related metric known as pointwise v-information gives information about the contribution from each datapoint in the dataset. However, these metrics suffer from robustness issues because they change too much from small perturbations in the datasets. This makes it hard to purposefully compare the difference in v-information between two datasets. My research is focused on creating a bound for this metric. I am using a method known as Bayesian Dropout to approximate the bounds of v-information. This will allow future researchers to use this metric with more confidence.

Conclusion

My STS research focuses on how biased datasets have an impact on machine learning models, whereas my technical research was about improving an existing metric that could allow researchers to pinpoint sources of bias. Completing these projects simultaneously gave me motivation because I was able to identify applications of my technical research. Ultimately, these projects improve the ethics of existing technology because they seek to improve the equality of service that technology provides.