Latent Probabilistic Model of News Sources

Project Alpha Prototype Report

Army Cyber Institute

Machine Learning for Media Bias

William Hiatt, Gabriel Matthew, and Deven Biehler

02/09/2023

TABLE OF CONTENTS

I. Introduction

This document serves as a guide to the Latent Probabilistic Model of News Sources. It will present new and updated details of the current state of the project as well as the direction the project is going. It will include information such as the social science aspects of the project, the open source models we will be using, models we will be creating, research that we will be using and research we will be doing internally, and any requirements that we currently have. Finally this document will quickly go over the bio's of group members on the team that are working on developing the model as well as all the stakeholders that the project will effect.

I.1. Project Introduction

The US Army Cyber Institute needs a model used for evaluating bias within individual news articles. This model will be used to give a bias rating on a specific news article. The model will grade the article on a scale based on the different bias's that the model looks for. With this scale the Army will be able to fight information warfare more efficiently, both on US based news sites and foreign news sites. The end goal of this project is to create a new model for the Army that will rate individual news articles on their bias and misinformation. This rating will be displayed in a way that is easy to read and follow. The model will use a mixture of currently available software as well as software created by the team.

I.2. Background and Related Work

Concerns about media bias and its potential impact on society have grown in recent years. Misinformation, polarization, and even discrimination can result from biased news. Journalists and media companies can become more aware of their own biases and work toward creating more objective and balanced news content by using a machine learning model that detects bias in news reports. Furthermore, news consumers can use such a model to critically evaluate the news they consume and make more informed decisions.

Due to large amounts of misleading information or complete misinformation, the US military has grown increasingly interested in fighting back. With their help, we are hoping to build a model to decrease the amount of misinformation and the speed at which it spreads.

There are some sites that currently provide a similar service such as allsides.com, mediabiasfactcheck.com, and adfontesmedia.com. These sites currently just look at the media outlet instead of breaking down individual articles. These outlets also tend to use humans to help determine this bias and that brings in another potential bias.

I.3. Project Overview

The media has a huge duty in the form of distributing news to the American public. The information plays a huge role in how people vote. The current offerings in finding media bias aren't as detailed as they should be and still contain a bias in themselves as it's often a human who is assigning these bias's. For example, allsides.com has CNN ranked as far left while adsfontesmedia.com has them as slightly left leaning. This is just a single example, but it shows that there is even a bias within deciding bias. In addition one post from a media outlet could contain no bias and be factually correct while another may contain a lot of bias and not be

factually correct. The goal of the team is to combat this by creating a model to find the bias, without any hint of human bias, of a single news article.

The team, with the help of mentors at the Army Cyber Institute, will create a new model that will address the issues presented in current media bias tools. Through research, designing, building, and testing the team will deliver a media bias tool to help fight information warfare.

The project is being built from the ground up. There are no limitations on how we create the model. The team has decided to use Python to create the model. Python is extremely popular for machine learning and will be a great platform to use. The code will be housed on GitHub, this allows for easy version control and collaboration.

The team is also in contact with two Social Scientists at Army Cyber Institute to help in finding the most important bias's to use as well as the best way to display the output. The team will be using sentiment analysis and fact selection as the primary drivers in deciding a bias. As the project progresses, we plan to include additional secondary drivers.

I.4. Client and Stakeholder Identification and Preferences

Our client is the US Army Cyber Institute with Senior Research Scientist Iain Cruickshank as our mentor and primary contact for the project. The product will be used and maintained by the Army. There are several stakeholders within the Army Cyber Institute including Iain and his colleagues.

II. Team Members - Bios and Project Roles

William Hiatt is a 4th year software engineering student at Washington State University. His skills include C#, Python, SQL, Java, C++, and GoLang. He has prior experience working as a software engineer intern at Kochava as well as a junior web developer at Washington State University. For this project, William will act as a team lead, main developer and designer of the model, and work with his team to create a viable machine learning model.

Gabriel Sams is a 4th year Computer Science major at Washington State University. His skills include C#, Python, Database development/management, SQL, machine learning, software development, agile process, test-driven development, and data science. His prior experience includes mobile application development, database development and management, and unit, end-to-end, and functional testing. For this project, Gabriel will act as a main developer and designer of the model, and work with his team to create a viable machine learning model.

Deven Biehler is a 4th year Computer Science major at Washington State University. His skills include C/C++, Python, SQL, machine learning, HTML/CSS, agile process, Haskell, and data science. For this project, Deven will act as a main developer and designer of the model, and work with his team to create a viable machine learning model.