Sentiment and Bias Analysis over Main-Stream News on Social Media

NLP Project Proposal

Mohammad Mahdi Abdollahpour

1 Overview

I will try to construct a dataset consisting of posts written by main-stream news channels such as CNN, BBC, Fox, Reuters, etc on social media platforms like Twitter, Facebook, Telegram, etc. Almost all of the data will be formal English.

2 Platforms

As the first attempt, I will start collecting data from Twitter (using official API) as I already have a developer account. Although due to sanctions there will be some difficulties, I will make a request to get a Facebook developer account too. The necessity and method of using other platforms and services will be examined in the future.

3 Channels

I will choose channels in a way that they cover the whole left-right political spectrum. The spectrum will be segmented into 5 zones: Far Left, Center Left, Neutral, Center Right, and Far Right. Then, each of these news channels will be put into one zone. It can be done simply by referring to websites that already classified news channels. But the process of assigning news channels to each zone and checking the validity of those websites needs further investigation.

4 Volume

If we assume that each channel publishes at least 35 posts per week and we can extract data of six months, we can gather an estimation of $35 * 4 * 6 = 840$ posts per channel. And if we assume that at least 3 channels for each zone will be present in our dataset we can have about $840 * 5 * 3 = 12600$ posts in total.

5 Challenges

Besides the fact that sanctions and other limitations can make data collection remarkably hard, it's worth mentioning another challenge that can potentially cause serious difficulties: Most news channels use social media platforms to publish just their headlines not the whole body of their news. Although using social media brings us the convenience of dealing with just a single data format, it can produce such challenges. If the results based on headlines are not satisfying, I may need to write a parser for each channel to extract the body of news from their original websites.

6 Motivations

I spend a considerable amount of my spare time reading the news. I enjoy it. I also believe that is it important for me to be aware of what happens around the world. Perhaps it is mainly because I believe that understanding the underlying causes of events (even shallow) in the world can be helpful and sometimes important in my decision makings. Hence, a dataset consisting of news covering most ranges of the political spectrum seems appealing to me. For average people, most of whom follow the news from social media, it can be enlightening to be informed about the biases present in the news. In other words, it can increase the chance of doubting and thinking more deeply about what is being told to them.

7 Major Tasks

Classification   All around the world, including the United States, most news agencies represent, implicitly or explicitly, different opinions in their news known as biases. Political bias is one of the major types of biases that can be easily found in the news. In this project, I will try to capture this type of biases to predict left vs right-wing bias in the news published by American news channels. I hypothesize that with relatively high accuracy (above 70 percent) it is possible to predict if a piece of news is from a leftist agency or a right-wing channel.

Sentiment Analysis   News agencies around the world publish news from different points of view about a single topic. Some of these points of view are radically different from each other. Sometimes even contradicting. For instance, we can observe that news published from channels that are in China or Russia about a topic like "human rights" are very different from news about the same topic publish by European or American agencies. I will analyze the sentiment of news published from agencies around the world about 5 to 10 (will be determined in the future) controversial topics. My hypothesis is that most western and eastern agencies are shifted toward opposite directions in most topics. Some topics that I propose are: "human rights", "immigration", "religion", "gender", "skin color", "government", "economy", "human labor", "global warming".