Feminist Text Analysis, Spring 2023

Can there be a feminist text analysis? Feminism, text, and analysis in a computational world

Response Blog Post – Week 9

AI Ethics: Identifying Your Ethnicity and Gender

This reading showcases the use of NLP and Machine Learning to build an AI model that can accurately identify gender based on a single tweet.

But it begins with the questions around the ethics of allowing AI to detect ethnicity and gender, followed up with more questions and then some more and before the author even tackles the pros and cons properly dives straight into building a model to detect gender from a tweet.

The article then showcases how that is achieved, how the data is cleaned and processed, how the model is built and finally applied to showcase its accuracy. However even the model itself leaves more questions than answers like what percentage of model accuracy is acceptable? Are tweets from celebrities the best training dataset to predict gender for everyone or just a particular niche of people?

But coming back to the ethical question of allowing AI to predict gender or ethnicity, it can be very unethical and harmful in many ways in my personal opinion and here is one example I had in mind.

Let's say you are applying for a job application with your resume and cover letter and the employers run it through an AI model. Now the AI model they have has been trained on their own dataset of employees and tries to detect those who are likely to be more "successful" at their jobs. Consider this to be in the finance industry which is dominated by mostly white men. Since, their AI model has been trained on a dataset that in itself has coded bias because it is based on a training dataset that itself is biased on both ethnicity and gender, this can cause a problem if somebody with a different set of ethnicity and gender applies (for example a non-binary Hispanic person). This is why I personally would not advocate using AI to detect ethnicity and gender unless the training dataset itself is adequately representative of a rather diverse demographic or there is an absolute need.

Beyond Fact-Checking: Lexical patterns as Lie Detectors in Donald Trump's Tweets

As the title suggests, this paper is on figuring out a model to detect lies in Donald Trump's tweets.

But why does that matter and why can't we simply label things as lies in a journalism context? Isn't fact just different from opinions / lies / misinformation? Because, as the author explains, "the line between factual claims and statements of opinion can be difficult to draw" (Graves, 2016, p. 92). There can be "opinionated factual statements" where the statement is indeed factual but draws on a substance of opinion, misinformation as questions where while you are directly not making a misinformed statement but ask a question that would lead to a thought of misinformation, and weaving misinformation into "otherwise truthful statements" (Clementson, 2016, p. 253) where the statement you are making it true but needs more context as without it misinformation is propagated.

What becomes an even greater difficulty is understanding not only whether a statement is fact or false but if it is false, was it an "honest mistake or strategic misrepresentation". There are debates on this questions itself whether and exactly when in journalism you should present something as an intentional lie.

The research paper's primary aim is to see if any language patterns emerge in Donald Trump's tweets to identify false information. It aims to do this by comparing the language of his tweets when he shares true and false information and find the lexical patterns in each to create a model for detection.

They focused on some patterns that are associated with lying in language like the use of words to express a negative emotion, certain pronouns etc. The study then found that indeed a lot of Trump's tweets that had false or misleading information actually did have a high number of such patterns existing. This can be used in journalism and scholarship beyond simple fact checking when needed.

Gendered Language in Teacher Reviews

This is an interactive visualization with a chart that allows exploration of words used to describe male and female teachers from a dataset of reviews from RateMyProfessor.com. A word or two-word phrase can be entered, and the chart generates to show the usage of words (per million words of text) broken down by gender (male & female) which can be further broken down into "positive" or "negative" ratings which enables further exploration to see in what context the word was being used.

So, basically when a word or two-word phrase is typed in, it crawls the database of reviews from RateMyProfessor.com, searches up how many times those words have been used per million words of text and breaks it down by gender and further breaking it down by positive or negative review and creates a chart from it while showing which department / subject the professor teaches.

I personally think the visualization does match the intention and carries out the task quite swiftly. The data visualized matches the question asked in at least its literal meaning.

For me I also particularly liked the break down by department as well. For example, searching for "programming" shows computer science followed by engineering as the ones for which the word shows up the highest

while “aesthetic” shows up the most for fine arts.

Also, it made me stop and think once that while we are trying to find out how certain words can have a gender bias in reviews, I as the prompter typing in words am making decisions on what words I am choosing that I think will have a gender bias (making me think of my own biases in gender language).

Film Dialogue

This project comes in a few parts to it.

The first analyses 30 Disney films and breaks down the screenplay by gender which showed that 22 out of 30 of those films have males dominating dialogues even in a film like Mulan where the leading character is female.

The second visualization contains a larger dataset of 2000 screenplays where the researchers’ linked characters with at least 100 words of dialogue to their respective IMDB pages which can showcase gender of the actor. This is then used to visualize screenplays on a scale of 100% of Words are Male to 100% of Words are Female and the chart is interactive so you can hover over one to see the name of the film and the breakdown of dialogue percentage by gender.

The third visualization takes into consideration only high-grossing films defined as films rankings in the top 2,500 of the US Box Office. It shows a general view of a scale from 100% male to 100% female like before. But it does a better job of showing just how less of the films have female dominated speech or even a 50-50 gender balance.

In this visualization, you can explore further by clicking on a film title and it shows the exact percentage breakdown of dialogue by gender, who the top 5 characters are (by number of words) and a breakdown of the male or female lines by minutes of the film from start to end showing the increase (or decrease) of each against the timeline of the film.

The fourth visualization shows a bar chart of the percent of dialogues by actor’s age broken down by gender. This shows that women actors have less dialogues available when they are over the age of 40 while for men it is the exact opposite, and they have more roles available with age.

The last visualization is a more interactive one where one can search up a particular film, see its gender dialogue breakdown and number of words by characters. There are also some filtering options like years and genres.

The visualizations do look really cool overall, and I just loved how clean and aesthetic everything is.

Gender Analyzer

This is a tool where one can enter certain text and it tries to analyze whether the text was written by a man or a woman. I frankly do not know how good this tool is and have been very confused as I have tried to put in text from woman writers which have been labeled as masculine and text from male writers labeled as feminine. It does say that it has an accuracy of about 70% which means it is not that perfect. But I just did not understand the use case or like why you even need this in the context it is supposed to be used.

This entry was posted in Posts on May 1, 2023 [https://femethods2023.commons.gc.cuny.edu/response-blog-post-week-9/] by Muhammad Rakibul Islam (Rakib).