Carnegie Mellon University

Closing Gender Gap: Analyzing the Pipeline Model Output and Building Extension That Detects Keywords

Ruiqi Wang Advisor: Haiyi Zhu

Introduction

Background

Previously, Ruotong Wang has developed a pipeline model that detects potential gender stereotypes and biases. A list of keywords was generated from the pipeline mode. My work focuses on the use the keyword list.

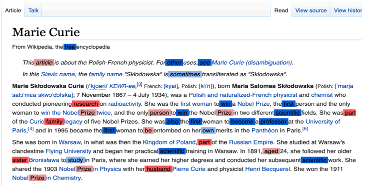
Abstract

In the first past, I developed a chrome extension that highlights the keywords on website as a visualizing tool. In the second part, I analyze how the keyword list overlaps with gender biases from previous literature.

Research

Chrome Extension

I developed a chrome extension that highlights all the keywords generated from the pipeline model. Red represents keywords that have negative coefficients, which are female-indicative, while blue is for male.



Article Talk Read View source View histor

Marie Curie

From Wikipedia, the free encyclopedia

This article is about the Polish-French physicist. For other uses, see Marie Curie (disambiguation).

In this Slavic name, the family name "Skłodowska" is sometimes transliterated as "Sklodowska".

Marie Skłodowska Curie (/ˈkjʊəri/ *KEWR-ee*,[3] French: [kyʁi], Polish: [kʲiˈri]), born **Maria Salomea Skłodowska** (Polish: [ˈmarja salɔˈmɛa skwɔˈdɔfska]; 7 November 1867 – 4 July 1934), was a Polish and naturalized-French physicist and chemist who conducted pioneering research on radioactivity. She was the first woman to win a Nobel Prize, the first person and the only woman to win the Nobel Prize twice, and the only person to win the Nobel Prize in two different scientific fields. She was part of the Curie family legacy of five Nobel Prizes. She was also the first woman to become a professor at the University of Paris,[4] and in 1995 became the first woman to be entombed on her own merits in the Panthéon in Paris.[5]

She was born in Warsaw, in what was then the Kingdom of Poland, part of the Russian Empire. She studied at Warsaw's clandestine Flying University and began her practical scientific training in Warsaw. In 1891, aged 24, she followed her older sister Bronisława to study in Paris, where she earned her higher degrees and conducted her subsequent scientific work. She shared the 1903 Nobel Prize in Physics with her husband Pierre Curie and physicist Henri Becquerel. She won the 1911 Nobel Prize in Chemistry.

Analysis of the PipelineModel Output

To quantitatively evaluate the result from the pipeline model, we perform the following evaluation: First, we identified a list of gender stereotype indicators from previous literature. Then, we used an automated method to determine how our keywords overlap with an existing stereotype.

Identify gender stereotype categorization lexicon in previous literature

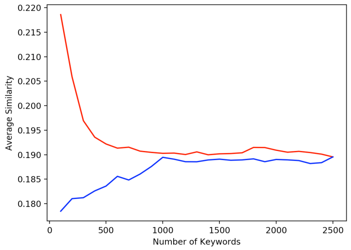
We manually collect 39 literatures on Google scholar using the keywords "Wikipedia gender stereotype", "gender bias", etc. Then, we list all possible types of bias and categorize them into eight dimensions ("powerful", "wealthy"," sporty", etc.).

Calculate the similarity between the SVM output and seed words

We quantify the similarity between our pipeline model output and seed words using the token similarity given by the word embedding model in SpaCy. We repeatedly carry out this calculation for the set of words with high absolute coefficients.

Result

The plot on the left captures the similarity between top keywords and the seed of appearance bias. The red line represents female, while the blue line represents male. The fact that the red line is above the blue line indicates that our female-indicative keywords are more related to appearance, which corresponds to the previous literature that people tend to focus on appearance when describing female.



Conclusion

Discussion

The chrome extension serves as a visualizing tool for the pipeline model, which helps people quickly capture the keywords that might indicate gender bias. The analysis of SVM output, which is based on SpaCy similarity, shows that the SVM model output fits with the gender bias from previous literature.

Future Work

Many interesting experiments have been left due to lack of time. Below are some ideas: (1). The chrome extension could be further developed by adding analysis to each highlighted word. For example, the bias categorization and SVM coefficient can be added; (2). It could be interesting to carry out interviews with Wikipedia editors and collect their feedback on this extension.

Acknowledgement

First, I would like to pay special regards to my supervisor, Prof. Haiyi Zhu, for her patient guidance in the weekly meeting. My special thanks are extended to Kai-Wei Chang, Steven Wu, Zachary Levonian, Caitlin O'Callaghan for their valuable and constructive suggestions during the weekly meeting. Then, I would like to thank Ruotong Wang, who leads this project, for her continuous support and encouragement throughout this semester.