

EECS 486 Project Final Paper: Gender Bias in Political Media

Katie Matton, Mara Gordon, Jennifer Winkler, and William Stager
University of Michigan

1 Problem Description

Gender bias in the media is a long-standing problem with harmful consequences. Popular news sources have considerable influence, and therefore, the context in which they describe women and the language they use when doing so has the power to significantly affect the public's perception of women. This issue was particularly relevant during the 2016 presidential election between Donald Trump and Hillary Clinton, and some have noted that biased reporting may have played a factor in determining its outcome. A report by Harvard University's Berkman Klein Center for Internet and Society found that mainstream media coverage during this time significantly favored Donald Trump. The study found that while the coverage on Hillary Clinton was largely focused on scandals, reporting on Trump's disreputable moments was balanced by discussion of his stances on serious issues such as immigration (Faris et. al, 2017). The potential for prejudiced language and reporting to negatively influence the way we perceive women, particularly in politics, acts as motivation for our investigation into the manner in which gender bias manifests itself in political media. Furthermore, a recognition of the extremity in the differences between viewpoints along opposing ends of the political spectrum has prompted us to explore whether there is a connection between political leaning and the presence of gender bias in reporting.

2 Project Goals

The primary goal of our project was to identify meaningful differences in the way women and

men are described in the political media. This includes an analysis of the specific words and contexts that accompany the media's discussion of members of each gender. As a secondary goal, we aimed to explore the relationship between political leaning and gender bias. Specifically, we wanted to examine whether patterns in the language used to describe men and women could be used as features to aid in the prediction of the political leaning of news sites.

3 Related Work

A significant body of research exists that analyzes the effect of gender on political opinions and voting. Many other studies examine instances of gender bias in the media and hone in on the differences in the way female and male candidates are described. By considering information gleaned from each of these areas, we can better design our project to identify patterns related to the intersection of gender bias and political preference.

In "Talking Politics on Twitter: Gender, Elections, and Social Networks," the researchers discuss whether well-known patterns of gender bias against females hold when looking at conversations on social media, and more specifically, on Twitter. The study looks more closely at relational power, a type of power that relates to personal relationships, and how the power shifts when the two candidates in an election are of opposite genders. For example, it looks at the adjectives used to describe male and female politicians on social media. The article concludes that when females run against males, their social media presence is more central to

their campaign and that there exists a differential distribution in power when discussing male and female candidates that favors males (McGregor and Mourão, 2016).

A research study titled "Framing the Fight" analyzes political media coverage of male and female candidates seeking state governor and U.S. senator positions. It finds that there are significant differences in the manner in which men and women are described even when competing for their own party's bid, and that this can translate into bias that harms female candidates during general elections (Bystrom et al., 2001).

In "Is She 'Man Enough'? Women Candidates, Executive Political Offices, and News Coverage", Meeks studies the news coverage of female candidates and their male competitors in multiple elections. She centers her research on the way that the candidates are discussed in reference to "feminine" and "masculine" political issues and character traits. She concludes that the coverage does differ across gender lines, especially in analysis of personal traits and stances on important matters. This was especially true for candidates running for White House positions (Meeks, 2012).

The study "Press Coverage of Mayoral Candidates," examines the role of gender bias in news reporting and the language used in discussing campaign issues during mayoral elections. The authors, Atkeson and Krebs, had seen many studies on gender bias in news coverage of national or statewide elections, where confounding variables such as party and competitiveness were apparent. They wanted to take a different approach to look at nonpartisan, open-seat mayoral races. The study found that the press coverage was not biased in favor of male candidates. Further, Atkeson and Krebs found that when there was a female candidate, local news coverage recognized mainly strengths of female candidates (Atkeson and Krebs, 2008). Our group thought this was an interesting study as it presented an opposing view to most of the other studies we had read.

One study we found examined the way that The New York Times portrayed Hillary Clinton and Sarah Palin in their respective elections. This study revealed that there were existing stereotypes in the news coverage of these

elections. The researchers who authored "All the Gender That's Fit to Print How the New York Times Covered Hillary Clinton and Sarah Palin in 2008," reported that the Times focused on female candidates' novelties while also emphasizing male candidates' masculinity. They also found that the Times gave more coverage to the male candidates than the female towards the end of the elections (Meeks, 2013).

Another article we analyzed looked at if there was a difference between liberal and conservative newspapers in coverage rates of female subjects. "Is There a Political Bias? A computational Analysis of Female Subjects' Coverage in Liberal and Conservative Newspapers," looked at the rate at which females were covered in 168 newspapers. The researchers expected conservative media to cover females less than liberal newspapers. However, there was no significant difference between the rates. The study did find that all newspapers covered males much more compared to females (Shor et al., 2014). This was a surprise to our group as it was to the authors of the article and we were curious to find if we would come to a similar conclusion.

A further literature review found that in addition to studies that confirm the existence and impact of gender bias in the media, there is also a large body of work that examines methods of analyzing text to identify instances of gender bias. "Detecting Gender Bias in the Coverage of Politicians in Irish Newspapers Using Automated Text Classification" is a thesis by Susan Leavy that thoroughly examines natural language processing and machine learning methods of identifying evidence of gender bias. Leavy found that the adjectives used to describe men and women were a particularly useful feature for recognizing instances of gender bias, which agrees with some of our findings from our project. She also found that verbs and single words were meaningful in discovering gender bias, which provides us with a possible area of exploration for future work. Other interesting findings from Leavy's work include that descriptions of female politicians were much more frequently accompanied by mentions of their spouses than were descriptions of males and that the specific policy issues associated with each politician had clear patterns along gender lines (Leavy, 2014).

These sources validate the importance of examining gender bias in the media and its effect on the ability of women to seek positions of political power. They also provide a reference for methods that may be useful in automatically identifying instances of gender bias. These existing works helped us produce additional insight into how gender bias manifests itself differently in sources with varying political leanings.

3 Datasets and Collection

Our dataset consisted of political news articles from online newspapers. We determined the news sources that would be represented for the liberal, neutral, and conservative sites, such as CNN, ABC News, and Fox News, respectively. We then crawled each of those sites to collect our data.

3.1 News Sources

To collect data for our project, we crawled 14 major news sources using a modified version of the web crawler we wrote for Assignment 3. We selected our news sources based on three major factors: the source's political leaning, how many political articles they contained, and how easy it was to crawl through only political content. Using a MarketWatch source, our team researched conservative, neutral and liberal news sources to crawl (Langlois, 2018). We filtered out any sites that contained only videos or made it hard to distinguish political content. From this, we picked five conservative sources, five liberal sources, and four neutral sources, as seen in Table 1 below:

Liberal	Neutral	Conservative
CNN	ABC News	Fox News
New York Times	PBS	The Hill
Politico	USA Today	Breitbart
Occupy Democrats	NBC News	Washington Examiner
Slate		Reason

Table 1: News Sources

3.2 Crawler

When crawling a site, we looked for links to articles that were from the same domain and were considered political content. To ensure an article was from the same domain, we utilized a similar methodology to that used in Assignment 3 which verified a page was from the "umich.eecs" domain. When attempting to confirm the article was political, we encountered a major issue. We were under the impression that we could simply search the html of every page for a flag to indicate if the content of a page is political. However, every site had a unique way to identify an article's content and searching through a page's entire HTML was cumbersome. To adjust, we chose our new sources carefully. More specifically, we selected sources that either contained exclusively political articles or had specific URLs which contained keywords relating to politics, such as "government", "election", or "politics". For example, when crawling Politico, we knew all articles were related to politics, so we could crawl the entire site freely. However, Fox News contains other content, such as business and lifestyle, so we checked if the URL for a particular article contained "politics". For each site, we collected the first 1,500 links that the crawler found and used these as our dataset. Once we gathered the URL list for each site, we proceeded to download each article. After accounting for 404 download errors, our final document collection consisted of 5,311 liberal articles, 4,134 neutral articles, and 4,880 conservative articles.

4 Approach

Our method was comprised of two major parts: feature extraction and the development of a classification model using a neural network.

4.1 Feature Extraction

Once the URLs had been collected, we classified each sentence in every article as either female or male. To determine the gender of a sentence, we looked for indicators, such as female and male pronouns or the name of an influential female or male politician. If there were indicators of both genders present in a sentence, the sentence was ignored and thus not placed in either set of sentences. For example, in Figure 1, the second sentence, “Trump claims that Comey made decisions thinking Hillary was going to win”, contains names of female and male politicians, which caused us to disregard this sentence in our analysis. However, the following sentence, “She passionately refutes this claim”, only contains a female pronoun, which classified it as a female sentence. We then extracted three sentence level features on these sets of female and male sentences.

Sentiment Analysis: First, we used the Python Natural Language Toolkit (NLTK) Vader sentiment analyzer to derive a sentiment value for each sentence. In order to translate the sentiment scores into document level features, we calculated the mean of the compound sentiment score of each sentence describing women and of each sentence describing men.

Adjective Analysis: Second, we used the NLTK’s part of speech tagger to identify adjectives and collected a count of each adjective used to characterize women and each adjective used to characterize men in each document. For example, in Figure 1, “untruthful” and “impartial” are counted as adjectives that describe men, while “passionately” and “malicious” are counted as adjectives that describe women. For each gender, we analyzed which of the adjectives were the most informative by computing the following probability-based measure:

$$\log \left(\frac{P(\text{gender} | \text{word})}{P(\text{gender})} \right) = \log(P(\text{word} | \text{gender}) / P(\text{word})) \quad (1)$$

To do this, we collected counts of all adjectives seen in female sentences and all adjectives seen in male sentences across all documents. $P(\text{word} | \text{gender})$ was then computed as the number of times that the specific word appeared in a sentence describing that gender. $P(\text{word})$ was calculated to be the number of times that word was used in a sentence describing either a male or a female. Instead of keeping all of the adjectives seen as features, we reduced our feature set to include only the the top 20 adjectives what we computed to be the most informative of the female gender and the top 20 that were the most informative of the male gender.

Linguistic Inquiry and Word Count (LIWC) Analysis: Lastly, we used the LIWC dictionary provided in Assignment 4 to count words in psychologically meaningful categories. For each word in each set of sentences, we searched the LIWC dictionary for that word. If present in the dictionary, the count of the associated psychological category was incremented. From the male sentences in Figure 1, “calls” corresponded to the “social” category, increasing the count in our LIWC dictionary.

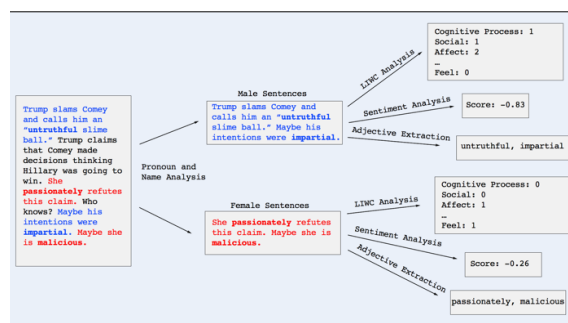


Figure 1: Feature Extraction Pipeline

The figure above shows the feature extraction pipeline, which begins with gender identification for each sentence. It shows each of the feature extraction methods for both the male and female sentences, which produced male and female feature sets for each article.

Once our feature set was extracted, we analyzed the data to find meaningful patterns in

features, such as which adjectives were most associated with either gender.

4.2 Classification Model

After collecting and storing the natural language features associated with each article in our dataset, we developed a classification model that uses them as input. The model is a neural network that outputs a prediction of the political leaning of a news article's source. We developed the neural network using the TensorFlow library and based its the core structure and implementation off of an example made available on GitHub (Khuc, 2018).

The input layer consists of one node for each feature value and one additional node for a bias term. The neural net includes a single hidden layer composed of 100 nodes. We tested using 50, 100, 150, and 200 nodes for this layer, finding that 100 achieved the highest accuracy. The output layer consists of a node for each political category, where the outputted value represents the probability of an article belonging to the category associated with that node. During forward propagation, we applied the sigmoid function (Equation 2) as an activation function for each hidden and output node. We used cross-entropy loss (Equation 3) as the measure the error of the output. This loss function was selected because it provides a more granular form to compute error when predicting classification labels (i.e. liberal, neutral, conservative) instead of numerical values.

$$\frac{1}{1+e^{-net}} \quad (2)$$

$$-\sum_{c=1}^3 y_{o,c} \log(p_{o,c}) \quad (3)$$

Where $p_{o,c}$ = predicted probability occurrence o is in class c

c = class (conservative, neutral or liberal)

$$y_{o,c} = \begin{cases} 1 & \text{If occurrence (article) } o \text{ is in class } c \\ 0 & \text{Otherwise} \end{cases}$$

5 Experiments and Results

To evaluate the performance of our classifier, we computed accuracy and precision and recall for each political category as metrics. We implemented 4-fold cross validation and averaged the results for each metric. Based on the distribution of articles across categories, our baseline accuracy is 37.08%. Our neural network model achieved 38.52% accuracy, which is slightly above the baseline. As this performance difference is marginal, it is unclear whether the natural language features we used provided any meaningful signal for the model to use. The precision and recall values for each political category are shown in the chart below.

	Precision	Recall
Liberal	0.44	0.40
Neutral	0.08	0.07
Conservative	0.38	0.60

Table 2: Precision and Recall Results

As displayed above, the precision and recall results were significantly lower for the neutral sites than the liberal and conservative sites. This is expected, as it is difficult to classify something that is "in the middle", as the patterns associated with it may not be as identifiable as something with a more extreme label. Neutral sites are likely to vary, with some leaning a little more towards the liberal side and some towards the conservative side, which may have added noise to our dataset, likely contributing to the relatively low accuracy of the system.

Due to the large size of our dataset, training the neural network took a significant amount of time, so it was difficult for us to test many different versions. We also think that with more time, we may have been able to find a neural network architecture that achieves slightly higher accuracy.

Some of the most interesting results of our project were captured as we analyzed and selected our feature set. The table below lists some of the adjectives that we computed to be the most informative as to the gender of a sentence.

One of the most perceptible patterns we observed was that adjectives related to aesthetic qualities, both positive and negative, were most frequently associated with women. These were also often of a sexual nature; “naked” and “sexual” are two examples. In contrast, adjectives accompanying the descriptions of men were frequently related to rationality, achievement, and respect.

Adjective	Female Value	Male Value
Gorgeous	1.79	-0.81
Obese	1.46	-0.45
Sexual	0.71	-0.12
Naked	1.13	-0.26
Assertive	1.10	-0.25
Successful	-0.20	0.02
Volatile	-1.36	0.08
Intellectual	-1.40	0.08
Factual	-1.10	0.07
Honorable	-1.18	0.07

Table 3: Most Informative Adjectives

6 Conclusions

Although our project did not definitively quantify the extent to which gender bias exists in the political media, it did reveal interesting patterns in the way in which the media describes women and men. One of our most notable findings was that adjectives related to appearance and sexuality were indicative of the subject of a sentence being female. We also found that adjectives related to intelligence, rationality, and achievement were more frequently associated with men than women. In future work, it would be interesting to examine the connection between political views and gender bias at the feature level. For example, an analysis of the frequency of the usage of biased adjectives on liberal and conservative news sites may help to reveal

whether gender bias is more embedded in the writing and thinking of either group.

Our examination into the connection between gender bias and political leaning was inconclusive. Our prediction model accuracy, which marginally exceeded the baseline value, hinted that there may be a correlation between the two variables, but without further analysis this cannot be stated outright. Our feature extraction generated results that we were able to associate with gender bias, but the features used as input to the prediction model were themselves not strict measures of gender bias. Therefore, any predictive power that the model had may not necessarily result from using gender bias as a feature, but may be related to patterns in the different political sites describe people in general. Another factor that may have contributed to the error in our model was the presence of quotes in the articles that we analyzed. If a site quotes the biased language of an individual in order to critique them, as was often done with Donald Trump quotes in the pre-election coverage, this may add noise to the results captured in our feature extraction process. Future investigation should consider removing quotes or handling them in a more precise manner.

7 Individual Contributions

Our team was extremely collaborative and worked very well together. Instead of dividing the project into independent parts for individuals to work on, we generally preferred to meet to work together. We found this was best as many aspects of this project introduced new challenges and ideas that we were able to work through faster as a group. Therefore, virtually all of the project was done as an entire team. However, each member did have a specific area of the project that they primarily led. These roles were as follows:

Will Stager: Crawler; Document parsing
Mara Gordon: Sentence gender identification; Sentiment analysis
Jennifer Winkler: LIWC analysis; Adjective extraction
Katie Matton: Neural network; Identification of informative adjectives

8 References

Dianne G. Bystrom, Terry A Robertson, Mary Christine Banwart. 2001. Framing the Fight: An Analysis of Media Coverage of Female and Male Candidates in Primary Races for Governor and U.S. Senate in 2000. *American Behavioral Scientist*, 44(12): 1999 - 2013. DOI: <https://doi.org/10.1177/00027640121958456>

Eran Shor, Arnout van de Rijt, Charles Ward, Saoussan Askar, and Steven Skiena. (2014), Is There a Political Bias? A Computational Analysis of Female Subjects' Coverage in Liberal and Conservative Newspapers. *Social Science Quarterly*, 95: 1213-1229. doi:[10.1111/ssqu.12091](https://doi.org/10.1111/ssqu.12091)

Lindsey Meeks. 2012. Is She “Man Enough”? Women Candidates, Executive Political Offices, and News Coverage. *Journal of Communication*. 62(1). DOI: 10.1111/j.1460-2466.2011.01621.x

Lindsey Meeks. 2013. All the Gender That’s Fit to Print How the New York Times Covered Hillary Clinton and Sarah Palin in 2008. *Journalism & Mass Communication Quarterly*. 90(3):520-539 DOI: 10.1177/1077699013493791

Lonna Rae Atkeson and Timothy B. Krebs. 2008. The Role of Gender in News Reporting and Campaign Issue Speech. 2008 *Political Research Quarterly*. 61 (2): 239 - 252. DOI = doi: 10.1177/1065912907308098

Robert M. Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election. *Berkman Klein Center for Internet & Society Research Paper*. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33759251>

Shannon C. McGregor and Rachel R. Mourão. 2016. Talking Politics on Twitter: Gender, Elections, and Social Networks. *Social Media + Society*, 2(3). DOI = <https://doi.org/10.1177/2056305116664218>

Shawn Langlois. 2018. How biased is your news source? You probably won’t agree with this chart. www.marketwatch.com. <https://www.marketwatch.com/story/how-biased-is-your-news-source-you-probably-wont-agree-with-this-chart-2018-02-28>

Susan Leavy. 2014. Detecting Gender Bias in the Coverage of Politicians in Irish Newspapers Using Automated Text Classification. School of Computer Science and Statistics, Trinity College Dublin. <https://www.scss.tcd.ie/publications/theses/phd/TCD-SCSS-PHD-2014-11.pdf>.

Vinh Khuc, Simple Feedforward Neural Network using TensorFlow (2018), <https://gist.github.com/vinhkhuc/e53a70f9e5c3f55852b0>