Does Gender Matter in the News?  
Detecting and Examining Gender Bias in the News  
Jamell Dacon¹ and Haochen Liu¹  
Data Science and Engineering Lab, Michigan State University

Introduction

To attract unsuspecting readers, news article headlines and abstracts are often written with speculative sentences or clauses. Male dominance in the news is very evident, whereas females are seen as "eye candy" or "inferior" and are underrepresented and under-examined within the same news categories as their male counterparts. In order to capture the attention of users, rich textual information such as news titles and abstracts present various forms of media biases such as ideological bias (i.e., biased articles that attempt to promote a particular opinion on a topic), coverage bias (i.e., media coverage regarding the visibility of topics or entities), selection bias, and presentation bias [1], thus contributing to the problem of gender bias. Since the 1950s, there have been studies on biased news reporting [2]. Media bias is both intentional as it reflects a conscious act and is sustained to present a systematic biased tendency[3].

Speculative Sentences/Clauses

Title example(s):
- “Women who want to succeed at work should shut up - while men who want the same should keep talking, research says”
- “Men have been promoted 3 times more than women during the pandemic, study finds”

Goals

• To detect and examine the phenomenon of implicit and explicit gender bias in the abstracts of news articles.
• To gain a sense of understanding of the gender representation in the news by examining the relationships between social hierarchies and news content.

Motivation

To identify how several forms of media bias (i.e., coverage bias, selection bias, and presentation bias) contribute to the problem of gender bias.

Datasets

- **MIND**: The MIND dataset was collected from the Microsoft News website, randomly sampled news for 6 weeks from October 12th to November 22nd, 2019. We obtained 96,112 abstracts.
- **NCD**: The NCD dataset was collected from HuffPost. The news articles were sampled from news headlines from the year 2012 to 2018 totaling in 202,372 news articles. We obtained 200,853 abstracts.

Analysis

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Abstracts</th>
<th>Category</th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIND</td>
<td>96,112</td>
<td>18</td>
<td>22,760</td>
<td>6,817</td>
</tr>
<tr>
<td>NCD</td>
<td>200,853</td>
<td>41</td>
<td>21,250</td>
<td>15,856</td>
</tr>
</tbody>
</table>

Table 1: Gender distribution test on the news datasets.

<table>
<thead>
<tr>
<th>Career Words</th>
<th>MIND</th>
<th>NCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spokes</td>
<td>192</td>
<td>121</td>
</tr>
<tr>
<td>Congress</td>
<td>191</td>
<td>49</td>
</tr>
<tr>
<td>Chair</td>
<td>255</td>
<td>20</td>
</tr>
<tr>
<td>Business</td>
<td>66</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2: Illustration of four intersecting career words (prefixes) across the two datasets for females compared to their respective male counterparts. The results are reported in terms of no. of gender-specific career words mentioned in each dataset per gender with their corresponding male/female suffixes.

References