Detecting Gender Bias in Job Descriptions: Machine Learning Final Project

Aimen Bajwa, Monica Kavathekar, and Jennifer Wright*
*George Washington University

***Abstract*—Science, technology, and engineering careers have long been dominated by men. This gender imbalance in STEM careers starts with the implicit biases that are placed upon our society and perpetuate the cycle of inequality. Small, almost unnoticeable biases in text online can discourage people from applying for a job that they are qualified for. Our project goal is to determine if the language used in the job description contributes to this gender imbalance. We will use Word2Vec word embeddings and the cosine similarities in order to identify bias in job descriptions that are posted online. We will take a sample of 100 different job descriptions from equal proportions of different industries in order to determine if the wording of the job descriptions contributes to the gender distribution within the field. *Index Terms*—Machine Learning; Gender Bias; Word2Vec; Cosine Similarities**

Your final project report will be in the following format.

Use the provided latex template and style files. The project is about both implementing and using. If you can make a case that your work is a substantially novel scientific contribution, it is OK to not implement the algorithms. Otherwise, you should be implementing the algorithms you want to use. Please check Final Project directions for potential kaggle contests, other machine learning problems, and Heilmeier's Catechism.

I. Final Project Proposal

We invite two-page final project proposals that give an overview of what you intend to present, including any results or conclusions you intend to share. We strive for engaging talks and focused discussions, and so proposals should display exciting ideas that can be communicated clearly and with brevity. A goal of the final project is to present talks that are informative, engaging, and even entertaining.

We encourage you to link to additional sources of your work (e.g., software, videos, websites, papers) within your proposal. We will strive to incorporate these additional sources into the review process, although full review of material beyond what is contained in the submission text is not guaranteed.

A proposal must include a title and a list of authors responsible for the work to be presented (one of whom must give the talk). It must be no more than two pages including references. It must be submitted as a PDF document, and we recommend that proposals use this LaTeX template. **You are allowed to form groups of at most four students.**

We seek submissions that are complete and concise. They should provide a full overview of the proposed final project. Your project proposal must detail the data that you plan to use, how you will pre-process it, and a precise plan of action, including what questions you would like to ask/problems to solve, machine learning algorithm(s) you hope to apply, how you will perform your evaluation (e.g., for supervised prediction you might use cross validation, looking at accuracy; then you might analyze your false positives/negatives to understand where and why the algorithms succeed/fail), a timeline for your work, and an explanation of what you expect to learn from your project.

This is meant to be open ended, and we don't expect any two projects to be similar. The goal here is for you to spend time thinking deeply about machine learning. To give you an idea of the scope, I am expecting you to spend 40 hours (per person) between now and the end of the semester on the project.

Final Project Proposal Evaluation

20 pts	Clearly state the machine learning problem
10 pts	Related work - Brief literature survey
10 pts	Data collection and/or dataset/corpus
10 pts	Data pre-processing
20 pts	Approach (Is this a novel application, are you implementing a novel method?)
20 pts	Evaluation (Depends on your problem but for example having ground truth data for evaluating accuracy in supervised learning is one example.)
10 pts	What are you trying to understand better to gain insights on a particular problem?

II. Introduction

It is well known that there is a large pay gap that perhaps comes from the gender gap in STEM occupations. Several factors have been proven to contribute to this, lack of exposure, lack of role models FIND MORE. In machine learning we have also discovered that the language we use is implicitly biased and contributes to systematic bias in society, we have also learned that it is possible to quantify the bias in our language by using word embeddings. The WEAT test and the WEFAT test performed by Professor Caliskan at George Washington University have proven that the same bias from the implicit association test can be found in the word embeddings we use to represent language. Knowing these two things we have decided it would be interesting to investigate whether the language that is used for the job descriptions contains bias and if so, does this bias contribute to the gender gap or reflect the the gender gap in STEM related occupations.

Science, technology, and engineering careers have long been dominated by men. This gender imbalance in STEM careers starts with the implicit biases that are placed upon our

society and perpetuate the cycle of inequality. Small, almost unnoticeable biases in text online can discourage people from applying for a job that they are qualified for. Our project goal is to determine if the language used in the job description contributes to this gender imbalance. We will use GloVe word embeddings and the WEFAT algorithm in order to identify bias in job descriptions that are posted online. We will take a sample of 100 different job descriptions from equal proportions of different industries in order to determine if the wording of the job descriptions contributes to the gender distribution within the field. We will display this information through a web plugin that operates as a Chrome extension.

III. Problem Statement

Determine if there is a link between the bias in the job description and the percentage of women who are in the specific occupation.

IV. Related Work

Bias in artificial intelligence is a new area. Just recently programmers have discovered that the data we use to train our algorithms also contributes to the amount of bias in the algorithms themselves.

In recent years gender bias in artificial intelligence has become a topic which is being discussed more and more in the tech world. Common AI systems such as Amazon?s Alexa and Apple?s Siri have been found to display gender bias at a massive scale with severe consequences. As this issue of gender bias is becoming more and more common, a lot of research is being done to identify and eliminate it. The January 2016 issue of the Journal of the American Medical Informatics Association, they talked about a study encompassing the evaluation of a system for automatically assessing bias in clinical trials.They used the Cochrane Risk of Bias (RoB) tool to define domains of bias and extract supporting text. Another research article which talks about identifying bias is the? Semantics derived automatically from language corpora contain human-like biases? written by Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan. They use the Word Embedding Association Test (WEAT) and the Word Embedding Factual Association Test (WEFAT) to evaluate bias in text.

You should properly cite the related work and methods. Enter latex bib entries to final_project.bib and use the \cite{bib entry keyword} command. One example is citing my bias in AI paper [**?**].

V. Dataset

The dataset that we found and will be using for this project consists of jobs currently posted on the City of New York?s official jobs website. The postings are updated on a weekly basis, so for our project, we will be using data from the week of November 25th, 2019. The dataset contains multiple thousands of jobs, so it is large enough to view a significant bias score. In addition, New York City was chosen as the sample area because the range of jobs is very wide-spanning from plumbing and maintenance work, to social media marketing jobs.

VI. Approach

In order to detect bias in job descriptions to better understand how it affects the actual percentage of women in a certain industry we have decided to use word embeddings. We have chosen to use Word2Vec, an unsupervised learning algorithm that produces dense vectors, that was trained by on Google news data. We made this choice because we thought the language used in news articles would be similar to the language used in job descriptions and could thus provide a more accurate sense of the similarity between words. In the word2vec library there are two ways to develop models, skip gram is the first and continuous bag of words is the second. We have decided to use a continuous bag of words in order to determine the similarity between words because if predicts the meaning of a word based on the words that surround it.

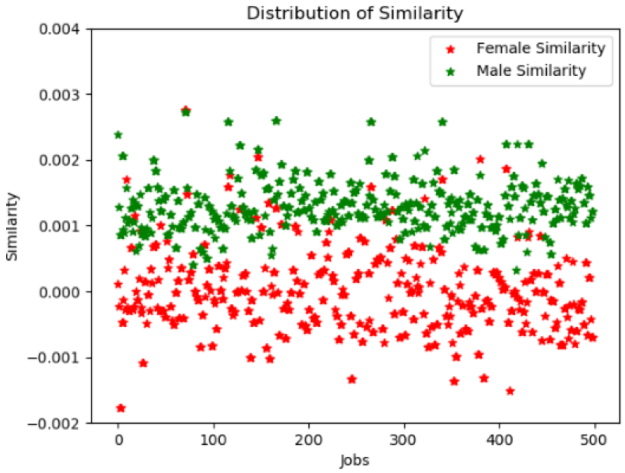
For the actual implementation, we will first parse the job description strings so that they are split up word by word, then we will begin to calculate the bias score. We plan to calculate the bias score by first comparing each word to words that are related to being female, we will take the average of how similar they are and use that as the female score. Then we will do the same for male, taking the words that are related to male and seeing how similar the job description is to them. We will be using cosine similarity to determine similarity between vectors as it is a well respected method. From there we will have a female and male score for each description and use this to determine if the job may attract more females or males due to biased language. Then we intend to compare the scores to other factors like average salary, industry and company in odrer to uncover it the language of the jab description has any correlation to the actual percentages in the industry.

VII. Experiments

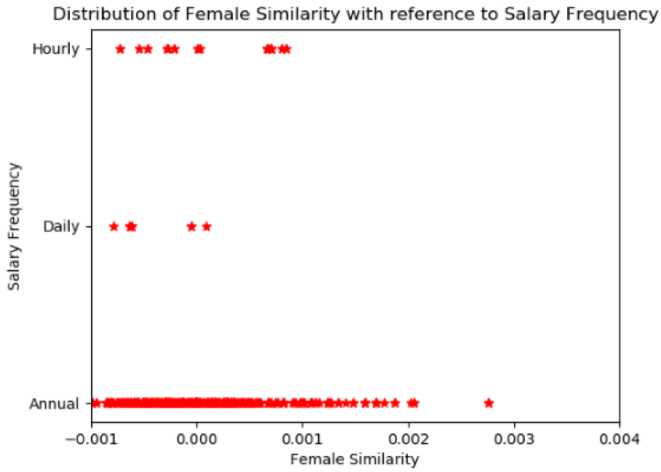
Our experiment involves comparing the cosine similarity to female and male based off of a list of words from Professor Caliskan to the type of job, industry and average salary. Our main goal is to discover if the actual percentage or women in the industry correlates to how similar the language was to being female. We also thought it would be interesting to look at salary and company to see any trends as well.

VIII. Results

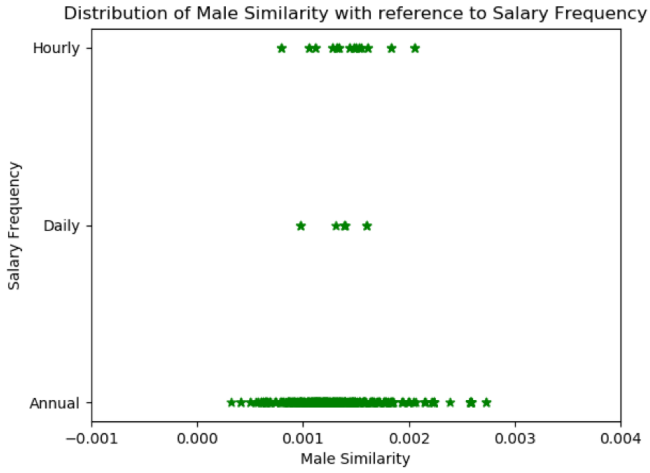
The results of the experiment were as predicted diverse. Some of the jobs were more similar to female language and some of the jobs more similar to the male words. This is as expected because some jobs target women and some men. We produced several scatterplots to display the results that were found and help us to understand the correlation.



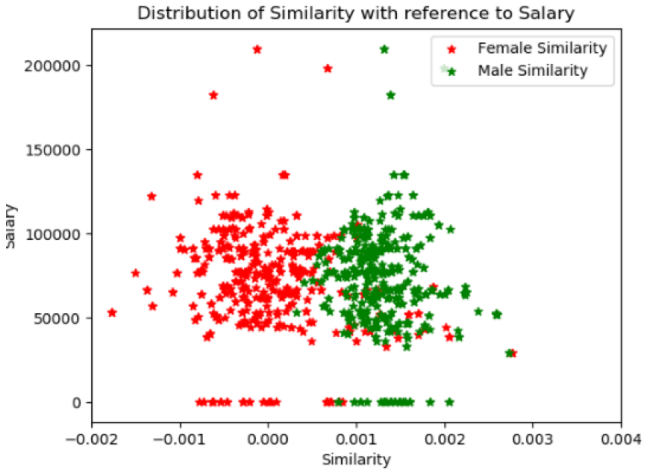
This plot shows the cosine similarity of the first five hundred job descriptions with relation to the list of male and female words. Here one can see that the descriptions are slightly more similar to male words than they are to female words. It can also be seen that while most of the male similarities are positive, about one half of the female cosine similarities are negative, meaning that they are much less similar while being lower. It is also interesting to observe that all of the cosine similarities are pretty low, less than 0.004 for the highest value, meaning that the results while interesting are not super powerful.



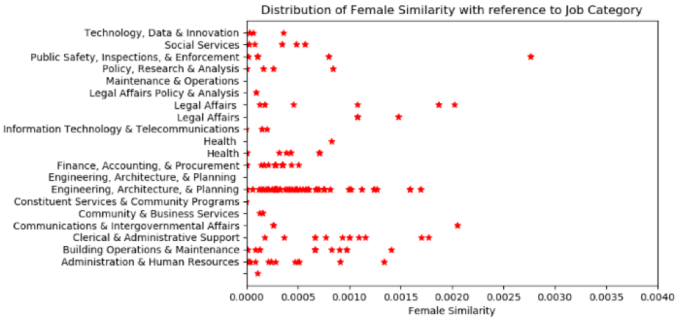
This graph shows the female cosine similarity versus type of salary. It shows that most of the jobs have an annual salary and just a few have daily or hourly. The daily salaried jobs seen to be on the lower end of cosine similarity in the range of the female cosine similarities and the hourly and annual salaries seem to cover the whole range of the female cosine similarities. The daily salaried jobs seem to be on the lower end in terms of similarity to female words.



For the male salaries, they are overwhelmingly annual, with very few daily and hourly jobs. It also appears that the jobs with the highest similarity to male are annualand the job descriptions with lower cosine similarity are more related to hourly and daily.

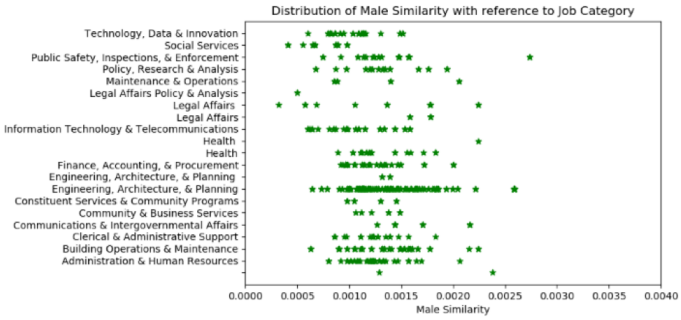


This graph shows the cosine similarity for men and women related to average salary. It appears that at all of the salary levels are significantly more similar to the male list of words than to the female list of words. It is important to note that the range of similarity is small so overall it is not that different.

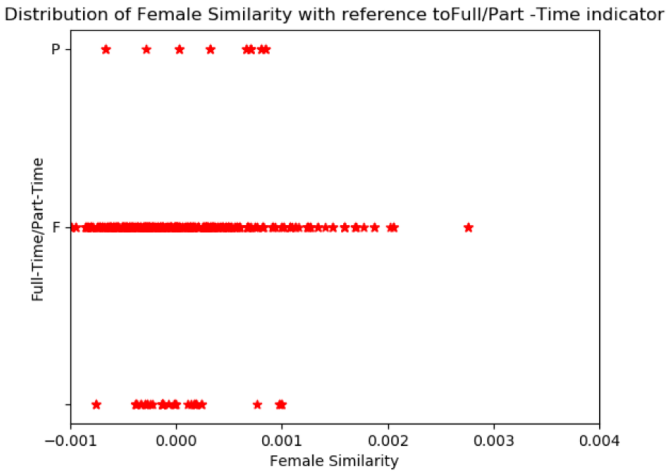


This job shows the categories of job types versus the similarity to female. It seems that there is a good range for each job

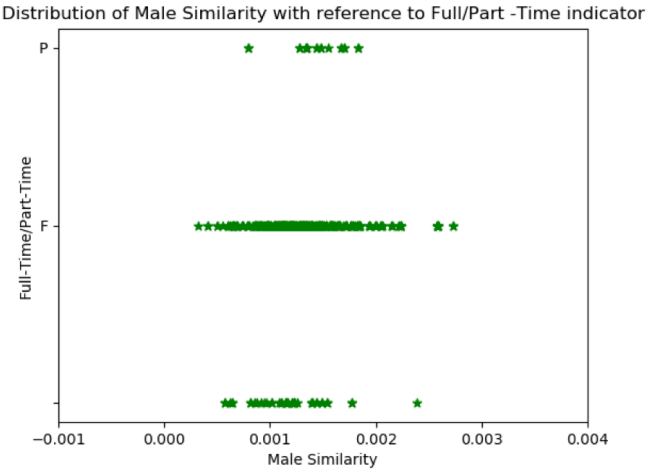
description. Communication, law, policy analysis all have high similarity to the female words. Public safety and enforcement has the highest similarity to the female words, this seems like an outlier and perhaps the description has gendered language.



This graph shows the job categories for the job descriptions versus the similarity to male words. It appears that each category has about the same range of similarities with relation to male words. The categories that have the most similarity are public safety and enforcement and engineering. This makes sense and lines up with the percentages in the workforce today. Other than that health and law also have a high similarity to the male words.



This graph shows the cosine similarity to females versus the type of job. It appears that overall the full time job section has a greater similarity to female and the part time category is sparse and on the lower side of the cosine similarity.



For this graph, the cosine similarity related to men is shown versus the type of job. The part time jobs are less similar to the male words and the full time jobs are overwhelmingly strongly similar to the male list of words. The uncategorized words seem to be few with less similarity to the male words.

IX. Discussion

The general trend seen across the similarity scores indicate that job descriptions show slightly more similarity towards male words as compared to female words in job descriptions. However, it was seen that in many instances both male and female shared the same similarities, and instances where cosine similarities were inconsistent between the two were not that significant in difference. In all job descriptions the maximum difference in male similarity and female similarity was up to 0.4 , less that 50% which is not that significant. The similarity distribution for each job description was viewed in comparison to the average salary earned, the salary frequency , is it full or part time and the industry the job belongs to. From the plots and analysis done, it was seen that the from these four job categories showed the most important results. From the plots it can be clearly seen that certain job industries such as Technology Data & Innovation, Engineering Planning & Architecture and Police, Research & Analysis have more male similarity as compared to female similarity shifting the graph to the right.

X. Conclusion

From this machine learning project, we can conclude that the overall level of bias in job descriptions is low. We found that the bias scores were mostly less than 50% for all test data, which means that the bias level is not significant. However, in the little bias that we did detect, there was more similarity to male words overall, meaning that more job descriptions are biased towards males. In addition, there were more STEM careers that showed a bias towards males. Conversely, careers in the social sciences were not significantly biased towards female words. Therefore, we can conclude that certain job industries do contain a small amount of bias.

[2] [1]

REFERENCES

[1] J. A. Hahm, "Attributing factors to the gender gap in stem education and a corrective measure to mitigate the shortage at a local level," 2015.
[2] C. Molly, "The gender gap in engineering," 2017.