Gendered Language on RateMyProfessor
486 Final Project Report

Amy Zhou, Elizabeth Dale, Kelly Henckel, Veronica Day

1 Project Description

Student reviews and ratings play a significant role in the career trajectories of university professors (MacNell, et. al 2015). Both historically and presently, instructors that come from underrepresented backgrounds have been found to receive lower ratings and gender-biased reviews. This perpetration of biases can pose a substantial threat to underrepresented instructors' careers in teaching and academia.

For our project we determined the genders of University of Michigan professors on RateMyProfessor through users comments and ratings. From this information, we found correlations between certain words and the gender of a professor. We predicted the gender of different professors by comparing the text of the comments of the other professors in our data set and their given genders. We attempted to specifically target adjectives and descriptive words, making sure to correlate professors not based on common words such as "lecture" and "class", or gendered words such as "she/he". In this way we were able to focus correlation more based on subjective description of a professor.

Our project was accomplished by crawling RateMyProfessor pages and collecting comments and tags on each professors page. We outline our method for obtaining the data for all the professors pages in our data collection description. From this data we identified the genders of professors using pronouns that commenters use (i.e he/she, him/her, etc.). If we were unable to find gender-descriptive pronouns, we used the professors' names to predict gender.

After we had a prediction of the gender of each professor we wanted to analyze the data we collected and see how differently female and male professors are rated as well as what language is used to describe different genders. By collecting and combining the the tags and words used in comments, we could correlate which words were most commonly used for each gender. We used a vector space model with TF-IDF weighting and the Nearest Neighbor (K of 1, 11, and 25) and Rocchio text classification methods to determine the gender of a professor.

To help showcase our findings, we created a web application where users can interact with the UI to search for a U of M professor by name, displaying the predicted and actual gender of that professor. The page also displays the top 10 most similar professors, the most frequent words used in that professor's reviews, as well as the most significant words related to gender prediction.

Our end goal was to see whether male and female professors are rated differently on RateMyProfessor and if there was enough difference in the language used to describe male and female professors to be able to determine their genders correctly, based on descriptive language used in student's comments.

Since student reviews are essential to a professor's career, it's important to recognize that bias may occur in these reviews based on professor demographics, such as gender. Specifically at the University of Michigan, while the EECS department is growing, the need for new professors is increasing. It is important to keep in mind possible bias that could manifest in these new professors' reviews, and the to keep this in mind when hiring new staff. Although RateMyProfessor is more often used in student's own decision making when determining whether or not to take a class, there is a probable significance between RMP ratings and the official reviews that the university collects each semester.

Our findings also support the importance of creating more awareness of bias in the workplace, specifically at colleges and universities. For example, why do words such as "organized" and "disorganized" occur more often for female professors than their male counterparts? What does this say about student's expectations from female versus male professors? Student's should be made aware of their own unconscious bias, in order to properly work against it.

2 Related Work

In 2015, researchers Lillian MacNell, Adam Driscoll, and Andrea N. Hunt found that students rated online instructors with a masculine name significantly higher than online instructors with a feminine name (MacNell, et. al 2015). Their findings held true regardless of the professor's actual gender; female professors posing as male professors received higher ratings than when they used their real name.

RateMyProfessor is not immune to these biases. In a project explored by Ben Schmidt (2015), 14 million textual reviews from RateMyProfessor were correlated with instructors' genders. Critical words, such as "mean", "harsh", and "tough", as well as soft words such as "nice", "kind", and "sweet", were unanimously correlated with female instructors. Conversely, words related to intelligence, including "smart", "brilliant", and "intelligent", were unanimously correlated with male instructors. As a result of his study, Schmidt found that female instructors were far more likely to receive feedback on their personalities, whereas male instructors were more likely to receive feedback related to their field of work or intelligence.

In an additional study of RateMyProfessor reviews, researchers found a strong, inverse correlation between reviews centered on instructor intelligence and the diversity of the instructors' field (Storage, et. al 2016). Departments that had high numbers of reviews containing "brilliant" and "genius" were found to have low numbers of female and African American P.h.D.s.

In a study published in the Journal of High Education (2016), researchers John A. Centra and Noreen B. Gaubatz found that students' reviews of instructors differed significantly depending on the students' genders (Centra, et. al. 2000). The study found that male students were more likely to rate female instructors lower in six of eight performance categories. Conversely, female students tended to rate female instructors higher, but not sufficiently high so as to offset the male students' ratings. There was no significant difference between student reviews of male instructors. This suggests that in majors with low female representation, such as engineering, female instructors may be rated lower.

Several studies have been able to accurately predict gender through subtle differences in language. In 2010, researchers Rob Thomson and Tamar Murachver were able to correctly ascertain the gender of authors of text messages in 91% of cases, despite removing any references to gender. Additionally, Cathy Zhang and Pengyu Zhang (2010) predicted the gender of bloggers using punctuation, sentence length, the parts of speech, and word topic. With information gain as a feature selection criteria and SVM (linear kernel) for classification, they predicted gender with 72% accuracy.

The correlations found suggest that preconceived beliefs of instructors' abilities based on gender and racial biases play a substantial role in students' reviews.

3 Data Collection

Crawler

We crawled the RateMyProfessor (RMP) online website in order to collect data on professor reviews, specifically professors at the University of Michigan, Ann Arbor. We started the crawler at the U of M "View All Professors" page on RMP (http://www.ratemyprofessors.com/search.jsp?queryBy=schoolId&schoolName=University+of+Michigan&schoolID=1258&queryoption=TEACHER).

This page contains urls that link to each individual professor's profile on RMP. (http://www.ratemyprofessors.com/ShowRatings.jsp?tid=1159280&showMyProfs=true). One issue we encountered was that these links were loaded dynamically using JavaScript, so simply downloading the html using requests.get(url).text would not reveal any of the links to professor profiles. In order to work around this constraint, we decided to use Selenium, a WebDriver application library. We used Selenium to automate opening a headfull browser in Google Chrome, and repeatedly scroll, find, and click the "Load More" button to dynamically load more professor urls onto the page. The html was then downloaded from the browser, containing the professor urls. One issue we ran into was making sure pop-up ads were manually closed before the program attempted to find and click the

“Load More” button, because a pop-up add would prevent the program from finding the button.

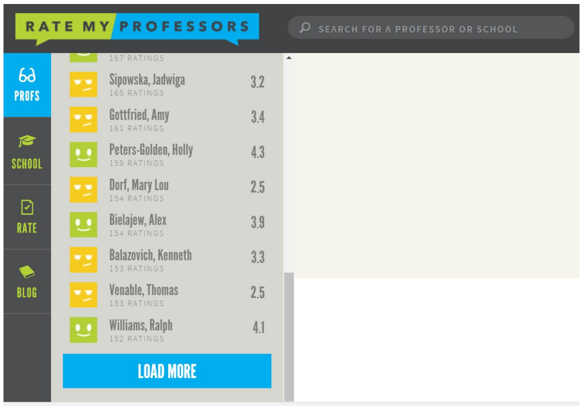

Figure 1: Sample “View All Professors” page for the University of Michigan

Parser

This html was then parsed for the professor urls, specifically pulling url’s that started with "/ShowRatings.jsp?tid=” and writing these urls to a file called professor_urls.txt.

The professor_urls.txt file was then used to open each of the professor links and download the html found on the professor’s RMP page, including overall rating, comments, and tags. We chose not to record information such as an individual post’s ratings and the chili pepper tag. We then used BeautifulSoup to find the correct divs related to the data we wanted to collect, and stored that data in a professor dictionary object.

We were interested in collecting:

1. Overall numerical ratings
2. Tags associated with the professor
3. Individual reviews written by students
4. Gender (derived in preprocessing)

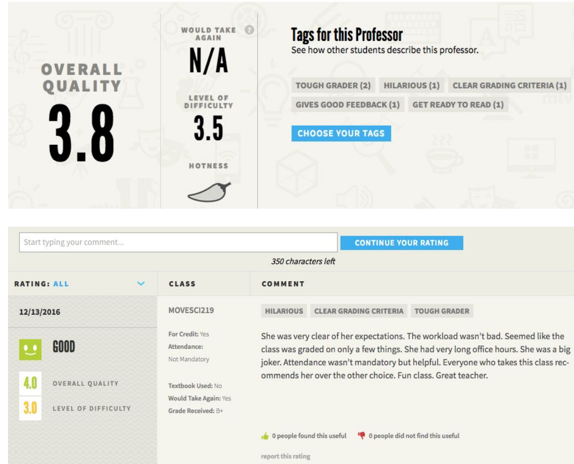

Figure 2. Sample RMP professor page

Each professor dictionary contained the professor ID (found in the url as “tid”), the full url path to the professor RMP profile, the overall rating or “score” that professor had, the professor’s name, a list of comments, and a list of tags and frequency of each tag. Each professor dictionary was written to a JSON file on disk, professors.json.


```
"2112983": {
    "url": "http://www.ratemyprofessors.com//ShowRatings.jsp?t
    "score": "3.8",
    "name": "Kathryn Clark",
    "comments": [
        "She was very clear of her expectations. The workload 
        She had very long office hours. She was a big joker. A
        class recommends her over the other choice. Fun class.
        "Dr. C was very passionate about her class. She makes 
        class requires a LOT of reading through scientific art
        and can be a tough grader but its only because she wan
    ],
    "tags": {
        "Hilarious ": "1",
        "Gives good feedback ": "1",
        "Tough Grader ": "2",
        "Clear grading criteria ": "1",
        "Get ready to read ": "1"
    }
}
```


Figure 3. Sample professor JSON dictionary.

The final item, the instructor’s gender, is not explicitly listed on RateMyProfessor. We ascertained the gender from the ratio of the number of ‘he’ pronouns to ‘she’ pronouns in the textual reviews during our preprocessing procedures.

4 Method Description

Preprocessing

The preprocessing of professor data begins by looking at the comments value from the JSON object passed by the Crawler. We considered 6 comments the baseline for sufficient data to create an inverted index. Thus, professors with 5 or less comments were not added to the inverted index .

For each professor that had more than 5 comments, a string was created by concatenating all retrieved comments. This professor comment string was then tokenized by splitting on whitespace and removing punctuation (commas, periods, semicolons, exclamation points, question marks). The most common English contractions were split into two different tokens. For instance, the word “she’ll” was split into “she” and “will”. These tokens were accumulated into a list.

This list was then passed to a function where certain stopwords (common words such as “the”, “a”, etc.) are removed. All gender pronouns (he/she, himself/herself, etc.) were also removed to ensure that none of the professor results contained any gender bias. To allow for more significant results, school-related terms that were commonly seen across comments (“lecture”, “exam”) were removed.

In order to identify if a professor was male or female, all instances of gender pronouns were accumulated in a dictionary where male pronouns mapped to a male pronoun counter and female pronouns mapped to female pronoun counter. A professor was labeled either male or female depending on whichever pronoun count was the highest. We never encounter an issue where the professor had an equal amount of pronouns for each gender.

After stopword removal and gender identification, term frequencies were calculated to create an inverted index. RateMyProfessor tags were also included as terms within the inverted index (e.g. the tag “Amazing lectures” was input as “_amazingLectures”). A dictionary of each professor’s name, gender, and term frequencies was written to a JSON file.

Determining Professor Gender

Taking the JSON object returned from the preprocessing step we used two different methods using our data to determine the gender of each professor. In both methods we use the Leave One Out method to compare a single professor(test) to the rest of the set(training).

The first method we used was Rocchio classification, where we combined all the the words in the text of male professors into one inverted index of term to professor ID, and all of the female professors into another. We then compared each of these male female vectors with a single professor vector from our testing set, using cosine similarity to see if they were more similar to the male vector or the female vector, and predicted their gender based on that result.

For the second method we used Nearest Neighbor to determine the results. We created term vectors for each professor in our training set individually (inverted indices of term to professor ID like the first method). We then compared each testing professor to all the other professors in our training set using cosine similarity. From the results, we used the top 1, 11, and 25 similar professors to determine the gender of a professor. For the top 1 result we just assigned each professor the gender of the most similar professor. For the top 11 and 25 similar professors we took the majority of what genders the top 11 and 25 genders were assigned to those professors originally and assigned the professor we were on that gender. (We originally used top 10, but realized that this may result in an even split of 5 male/5 female professor vectors, and switched to using K = 11 instead).

Once our results were calculated for each professor, the results were stored in a JSON object, written to disk, to be used in the final frontend interface.

5 Experiments and Results

Visualization of Results

To make our results publicly accessible, we developed a front-end visual titled *Gendered Language in RateMyProfessor*. The website dynamically reads in the JSON files created by our methods, interprets the data, and presents results for individual professors. Users can search for any professor in our database and view statistics, including the RateMyProfessor rating, actual gender, predicted gender, most similar professors, and most common words.

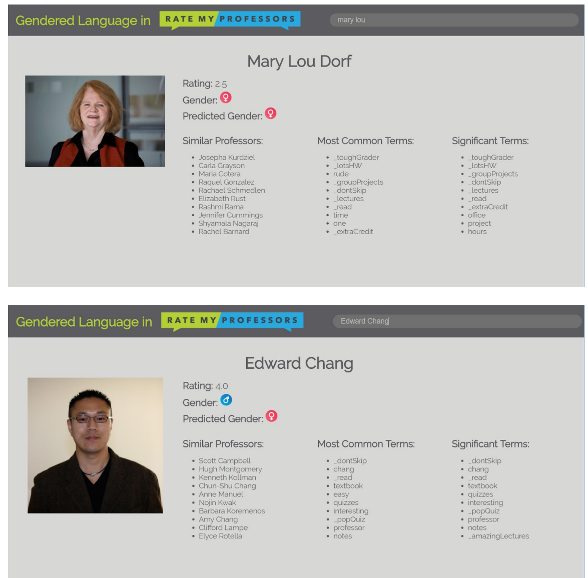

Figure 4. Sample professor pages.

When the website is initially opened, the three professor JSON files (professors.json, profQueries.json, and profTerms.json) are parsed using JQuery. The professor information from each file is combined into one hash indexed by ProfessorID. A second index is also created, mapping professor name to index.

When a user searches for a professor, the professor name hash is searched for exact substring matches. When there is a match, the matched professor's statistics are loaded from the professor information hash. The matched professor's image is also pulled from the Google API using a Custom Search Engine image query for "professor name" +

"umich" + "professor". The first image from the umich.edu domain is pulled, scaled, and displayed on the page.

Finally, the user can navigate to a basic homepage containing information on our methods and results by clicking on the page heading. The home page is also displayed by default when the user inputs an invalid professor.

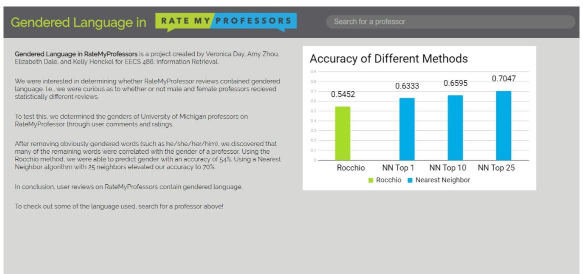

Figure 5. Homepage containing the project description, methods, and results.

The tool was created using the Vue.js framework in conjunction with HTML, CSS, and JQuery. In the version deployed to our Github, the JSON files were generated using the classifier with the greatest accuracy: Nearest Neighbor - 25.

6 Conclusions

Results (Using sample of 420 professors)

	<u>Accuracy</u>
Rocchio	54.23%
K = 1 Nearest Neighbor	56.28%
K = 11 Nearest Neighbor	61.87%
K = 25 Nearest Neighbor	70.24%

From the different methods we used to determine the gender of a professor the Nearest Neighbor method with different numbers (1, 11, 25) out performed the Rocchio method (see Results above).

For the top words used in reviews for male and female professors there was some overlap, but there were also many terms that were unique to each set, showing that in general there is some difference in the terminology used to describe female and male professors. Based on our results, no matter which words were used in the reviews and tags in professor comments, we can conclude that there was enough difference in the words used to describe male and female professors (with pronouns removed from the text) to be able to determine the gender of a professor.

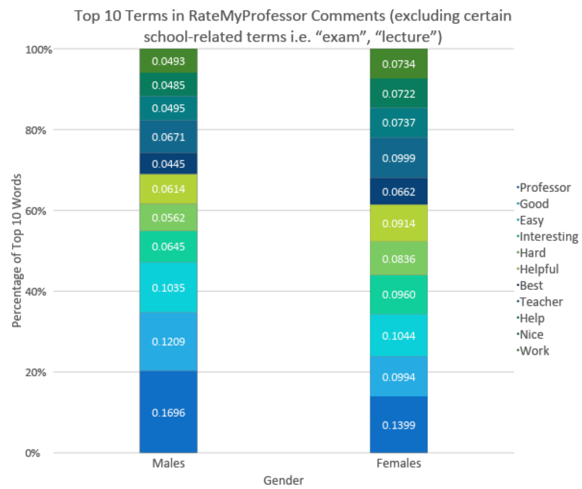

Figure 6: Top words and percentages for each gender.

We did not remove gendered words besides pronouns "he/she" etc. and could have benefited from removing more terms such as "mother" or "woman".

In terms of expansion, we would have liked to group descriptive words into categories of terms to see if certain categories are used more often for one gender over the other. For example, words describing intelligence ("brilliant", "genius", "smart") may have been used more often with one gender than another. We could have used an LDA topic modeling to accomplish this. We also would have liked to examine review ratings in correlation with the text reviews. This would have given more context to the actual text review for each professor. We would have also been interested in seeing if there was a way to know the gender of the reviewer, and the discipline the professor teaches in, to see if either of these factors affects a bias in professor reviews.

7 Individual Contributions

Veronica: Responsible for the RateMyProfessor web crawler/scraper for collecting and parsing data. In the final report, she was responsible for the Related Work and Data Collection sections.

Amy: Responsible for preprocessing text and determining professor gender. In the final report, she was responsible for the Preprocessing section.

Elizabeth: Responsible for professor queries and calculations. In the final report, she was responsible for the Project Description, Calculations, Conclusion sections.

Kelly: Responsible for frontend design and result visualization. In the final report, she was responsible for the Related Work and Experiments and Results sections.

We believe each group member contributed equally to the overall project, poster presentation, and final report.

8 References

Storage D., Horne Z., Cimpian A., & Leslie S-J. (2016) The Frequency of "Brilliant" and "Genius" in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields. PLoS ONE 11(3): e0150194. doi:10.1371/journal.pone.0150194

Zhang, C., & Zhang, P. (2010). Predicting gender from blog posts. University of Massachusetts Amherst, USA.

MacNell, L., Driscoll, A. & Hunt, A.N. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching
Innov High Educ (2015) 40: 291.
https://doi.org/10.1007/s10755-014-9313-4

Schmidt, B. (2015) A public exploratory data analysis of gender bias in teaching evaluations. IEEE Conference on Visualization (2016).

Centra, J.A. & Gaubatz, N.B. (2000) Is There Gender Bias in Student Evaluations of Teaching? The Journal of Higher Education (2016) 71:1.
doi:10.1080/00221546.2000.11780814

Thomson, R. and Murachver, T. (2001), Predicting gender from electronic discourse. British Journal of Social Psychology, 40: 193-208. doi:10.1348/014466601164812

Selenium Browser Automation, docs and version 3.11.0 available at
https://www.seleniumhq.org/