A Catchy Title about Ladies in Movies

Subtitle: Dr Grace Augustine is Our Hero

By Michelle Kim, I-Ching Wang, and Nikki Haas

**Abstract**

Is totally written last. It’ll probably be a paragraph, so let’s fill up a paragraph. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl. All work and no play makes nikki a dull girl.

***Intro***

For young girls growing up in our society, there are few things more disenfranchising than going to the movies and watching a hapless princess wait for others to act in her place.  Some women in film will be given a voice of their own, but too often men are given the responsibility of conveying the important ideas and plans.  At least, that is how it feels, but is it really true?  For our project, we want to test this hypothesis.  We will use natural language processing and deep learning via convolutional neural nets to evaluate screenplays for hundreds of popular movies from many genres over a period of several years and compare the complexity of scripted lines for male and female roles.

This project is important because young girls would benefit from seeing films that more accurately reflect the various roles women play in society as leaders and innovators.  Not only is length of dialogue important, but the content.  There have been many instances where women roles are given fewer lines, and in addition, the majority of their lines were about their male leads instead of anything substantial that contributes to the plot.  If we can prove that popular movies are following the trope of simplifying women using stringent statistical methods, perhaps we can help Hollywood see its limitations and effect change.

***Background***

In 1985, the cartoonist Alison Bechdel created the Bechdel test for works of fiction. This test checks to see if a) two women in a movie, television show, play, or novel b) speak to each other c) about something other than a male character. Since 1985, this test has been used for online resources, movie critics, and the media to judge the entertainment industry’s representation of women.



Figure : Bechdel Test by Alison Bechdel

Another paragraph something about automating the test using neural nets and NLP. Talk a bit more about why our approach is a good facsimile of a Bechdel test.

There are already various studies that focus on the length of screen time characters have based on gender, as well as how much revenue and ratings those types of movies get. We’ve taken a look at some of those studies and found that many times, women had far less screen time than men, even if the storyline has them listed as the main character. However, movies where screen times between the two genders had less of a difference seemed to perform better when they did exist. None of these studies went into an exploration of the actual content of the dialog that was spoken. Our study confirms a lot of the trends that these previous studies have shown, and will delve deeper into the differences on complexity of the dialogue between men and women as well as develop a prediction aspect based on the content.

Because there are known differences in the dialogue of men vs women in scripts, we are interested in seeing if these differences widen across genres of movies. By implementing NLP algorithms to process scripts, nature of the line being spoken, and looking at the Bechdel test, we aim to see if our model can properly identify if a line belongs to a man or woman.

***Methods***

*Data Collection*

Our model uses the Film Corpus 2.0 which is a collection of movies from various genres up to the year 2015.

From this Film Corpus, we took all the dialog files and parsed out information about the title of the movie, the genre, and individual lines. Each line represents roughly 10 second of on screen time. We have written a python pre-processing script that matches the lines with the character name that is speaking it. Most of the cues such as Cut to, exit stage, etc have been removed as much as possible. although since each script is written a little differently, there are still some lingering characters in some places. In addition, we have only kept characters that had a total of 20 or more lines. Characters that had too few lines didn’t have much of a vocab was deemed as not important enough to the story line.

We then randomly selected roughly 1000 characters to characterize the gender. This was based on the actor or actress that represented the character, though in few cases, there may be a mismatch. This is more common with LGBT characters or characters in animation, where a character may be voiced by an actor or actress of the opposite gender. The rest of the characters were used to build and test our model.

***Results and Discussion***

*Exploratory Analysis*

*Modelling*

Our first thought was to see how accurate our model would be by calling every character male. From our random selection, had already seen quite a heavy skew towards male characters, as shown in the exploratory analysis. By assigning “male” to every prediction, we found that this already gave us a ??% accuracy in calling the correct gender of the character, without looking at any contextual data.

Our first model we tried to use was a logarithmic approach. This was mostly to have a baseline for our study to improve on. Outright, we can see the approach perform poorly in

We should also talk about the issues with the data here – how there are much more data points for male than there are female and that the amount of speaking time for women is significantly less, so there’s less training material.

*Things we’re gonna do*

Bag of words first as a baseline

Talk about extending to CNN and why that’s so great.

*Statistical Inference*

Now with fancy pictures!



This picture is a place holder for the vocab size based on gender. I assume we’ll find that the mean for men is higher than the mean for women, and that the p-value will show that men really have more words than women. Probably do something similar with line count.

*Logistic Regression*

We’ll dump the vocab into a one hot encoder, and see if logistic regression can identify male/female.



placeholder

If it can’t, we’ll use PCA to look at genre, vocab size, and vocab to see if there is much of a difference.



placeholder

***Next Steps***

We’re gonna use Keras and CNN, because nobody uses straight Tensorflow anymore and we want to use this paper when applying for high paying jobs. More amazing pictures using Seaborne.

***Resources***

Bechdel, Allison. Bechdel Test. <http://dykestowatchoutfor.com/>

Bechdel Test Movie List: <https://bechdeltest.com>

Kim, Yoon. Convolutional Neural Nets for Sentence Classification. <http://www.aclweb.org/anthology/D14-1181>

Anderson, Hannah and Daniels, Matt. Film Dialogue <https://pudding.cool/2017/03/film-dialogue/>

Geena Davis Foundation. The Reel Truth: Women Aren’t Seen or Heard. https://seejane.org/research-informs-empowers/data/

*Links to data sources*

UC Santa Cruz. Film Corpus 2.0 <https://nlds.soe.ucsc.edu/fc2>

Polygraph’s Film Dialogue Set https://github.com/matthewfdaniels/scripts

*Link to our github*

<https://github.com/mathaholic/w266_final_project>

*Contact the Authors*

[icwang@berkeley.edu](mailto:icwang@berkeley.edu)

[mathaholic@berkeley.edu](mailto:mathaholic@berkeley.edu)

[michelle.m.kim@ischool.berkeley.edu](mailto:michelle.m.kim@ischool.berkeley.edu)