Analysis of female character roles in movies through dialogues

Amal Alabdulkarim
Department of Computer Science
Columbia University
New York, 10027, NY
aa4235@columbia.edu

Jing Qian
Department of Computer Science
Columbia University
New York, 10027, NY
jq2282@columbia.edu

1 Introduction

Movie dialogues are a projection of peoples inter-actions in real life. We can use these dialogues to understand the actual human interactions and get a hint of the trend of thought and its change through time. In this project, we want to understand the female character roles in movies through studying these dialogues. We are going to analyze the evolution of the dialogues through time, genre and rating of movies.

Researchers studying linguistic and social phenomena sometimes use fictional data. In (Danescu-Niculescu-Mizil and Lee, 2011), researchers have shown that studying linguistic convergence can be done by analyzing a movie dialogues corpus. They created the Cornell movie dialogues corpus to answer their research question; Has linguistic convergence become so deeply embedded in the language-generation process as to become an unconscious response? They hypothesize that fictional dialogues offer a way to study this question since authors create the conversations but do not receive the social bene-fits (instead, the imagined characters do). They used two methods to analyze the data, linguistic style by using LIWCs nine category function words and convergence measure to measure the occurrence of these linguistic style features in dialogues. The results of this research showed significance coordination between characters and their use of function words. Also, it showed that female characters (F) are more influential than male characters (M) in their dialogues and tend to trigger more linguistic convergence with other characters. Convergence in F-F dialogues is more significant than F-M dialogues and M-M dialogues which has the lowest convergence. This last unintentional finding suggests that this corpus has more to offer regarding gender roles in movie dialogue.

In (Skowron et al., 2016), the authors proposed an integrated approach to detect character types from movie dialogues. Based on a merged data set of action movie corpus, they designed experiments to compare different models (with different feature representations) by their capacities to identify character types, including several models with representations of utterances, graph models and their Graph+N-grams model. Their integrated model has the best performance in overall and four distinct dramatic character type detection. They also did single-movie analysis in depth and supported the feasibility to profile one characters verbal behavior. Moreover, the change of ones verbal behavior implies the development of that character.

In the literature, several studies focus on understanding gender roles in media, especially in movies and television shows. (Movies, 2018) Analyzes gender in two superhero movies Wonder Woman and Man of Steel. The researchers used five linguistic features, amount of speech, interruptions, questions, minimal responses and hedges. Their findings showed a correlation in three of those features (amount of speech, interruptions, and questions) and no clear patterns in the other two features. More specifically, the findings showed that men speak and interrupt more than women, while women ask more ques-tions than men.

In (Painter and Ferrucci, 2017), the researchers studied female journalists characters roles the television show House of Cards. They found that the female journalists characters were not de-picted negatively, but at the same time, they were

not presented as positively as their male peers. The researchers also analyzed other aspects of the female journalists' portrayal in this television show such as their ethical implications which might affect how society sees female journalists negatively.

In (DeTardo-Bora, 2009), researchers examine how female criminal justice professionals are portrayed in fiction. The researchers analyzed ten prime-time television show and observed sixty-nine characters, their findings showed an over-representation of female criminal justice professionals but in subordinate roles and with emphasis on them being sexually attractive. Moreover, similar to the research on female journalists, the characters were not depicted negatively but were not portrayed as good as their male peers.

(Glascock, 2001) Conducted a study on prime-time television shows on four of the most popular networks among the audience during prime-time according to Peoples Choice. Their primary goal is to determine how the roles of female and male characters have changed on prime-time television since the 1970s. Their first research questions focused on the change of the demographics of these movies. To answer this question, they measured several demographic features. Examples of these features are the number of female/male characters, marital status, age, and appearance features. Their second research question focuses on the differences in the behavior of female/male characters and how it changed in recent years. Their general hypothesis is that the female role in movies has improved through time, and they measured this using their demographic and behavioral data. Their findings showed that there was a change in how female characters are depicted in movies. However, female characters are still underrepresented in movies in general when compared to their male counterparts. To answer these questions, the researchers formed a committee of five coders. The coders' task was to record the demographic and behavioral features of each movie according to a coding scheme that they initially agree on after analyzing a subset of the data. The researchers used Scotts Pi to assess the inter-coder reliability and also used chi-square correlation as their primary test to analyze demographics. Their findings exhibited significant gender differences

in several features, such as marital and parental status as female characters are usually assumed to be married or with children but that was not the case for male characters who were usually single. Also, even though more females characters now have jobs in comparison to older movies when they were mostly unemployed. Their jobs are generally in lower paying and less prestigious occupations. The finding supports this claim, as the male characters were twice as likely to be bosses in movies as female characters. With regards to physical appearance, female characters usually stay younger, more attractively dressed, and more likely to have red or blonde hair than male characters. On the behavioral side, female characters were more verbally aggressive than males while male characters were found more physically aggressive than females. While the gender gap on television has shrunk, the change has been slower than what the researchers initially expected. One reason could be that some of these discrepancies are merely reflective of the real world. Another reason may be innate in the movie industry, which the researchers attributed to the relations between the off-screen representation the and on-screen demographics as they showed evidence of male-dominance in the industry in their findings.

In (Neville and Anastasio, 2018), the authors tried to find whether the gender and age inequalities in the U.S. movies found in 2005 by Lauzen and Dozier have changed in 2016. Following the methodology of Lauzen and Dozier, they coded characters in the 50 best-selling U.S. films manually for gender, age, leadership status and etc. As an extension, they also introduced film genre, levels of aggression into coding. They found that: although female major characters increased significantly, male characters still dominated. The age inequality in both men and women decreased and the decrease was more significant in mens. There is no significant gender gap in leadership roles, social aggression and holding goals. The authors called for more female and older characters in films, which could promote gender and age equality in real life.

One method for analyzing gender roles in movies is the Bechdel test, which was introduced by Alison Bechdel in her comic Dykes to Watch

Out For (Bechdel, 1986). The test aims to measure womens representation in fiction, and it consists of three questions.

1. Does the movie have at least two women in it?

2. Do they talk to each other?

3. Do they talk to each other about something besides a man?

In (Agarwal and Dey, 2015), the authors aimed to automate the Bechdel test through analyzing each question separately and then combine these features into a support vector machine model to predict if these movies passed the test or not. They relied on multiple sources of data such as the labeled Bechdel test dataset and IMDB dataset. Identifying features were not trivial; there were some issues in identifying character names; certain characters are credited with a name different from the way they appear in the screenplay. This issue may have caused several false negatives in answering the first question and second question. For the third question, the researchers combined several feature vectors, such as topic and linguistic features. The features that best answered this question were network analysis features. In general, their combined task achieved 80% F1-score. Their results showed that features based on social network analysis metrics (such as betweenness centrality) are most active. As they noticed in movies that failed the test, that the female characters were less centralized when compared to the movies that passed. Word unigrams, topic modeling features, and features that capture mentions of men in conversations are less effective in their experiments. They attributed it to a result they found in their analysis that even in conversations where men are not the central topic there may be a mention of a man. Through the analysis of the results, the researchers were able to show a significant correlation between the importance of roles of women in movies with the Bechdel test.

We are specifically interested in understanding female character roles in movies and how it changes through time, movie genre and ratings. Besides, we want to find answers to the following questions:

1. How do female character topics of dialogues change with time?

2. Does the dialogue of the female characters differ across genres?

3. Does the number of female lead roles versus male lead roles differ across genre?

4. Do movies with higher ratings tend to have male-leading roles than female-leading roles?

We anticipate that female characters dialogue topics became more diverse in recent movies compared to older ones, this can be measured using Linguistic Inquiry and Word Count (LIWC) to conduct a Bechdel test on the dialogue level and Latent We also think that female characters have a more powerful appearance in specific genres. Which we will measure using the number of lines, and the significance of their roles in those genres. Lastly, we expect that movies with higher rating tend to have a stronger male presence.

References

Apoorv Agarwal and Shirin Ann Dey. 2015. Automating the Bechdel test. pages 830–840.

A Bechdel. 1986. *Dykes to Watch Out for*. Dykes to Watch Out for Series. Firebrand Books.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs.

Kimberly DeTardo-Bora. 2009. Criminal Justice Hollywood Style: How Women in Criminal Justice Professions Are Depicted in Prime-Time Crime Dramas. *Women & Criminal Justice*, 19(2):153–168.

Jack Glascock. 2001. Gender Roles on Prime-Time Network Television. 8151.

Contemporary Superhero Movies. 2018. The Superman Speaks and the Wonder Woman Keeps Quiet :. pages 61–90.

Conor Neville and Phyllis Anastasio. 2018. Fewer, Younger, but Increasingly Powerful: How Portrayals of Women, Age, and Power Have Changed from 2002 to 2016 in the 50 Top-Grossing U.S. Films. *Sex Roles*, pages 503–514.

Chad Painter and Patrick Ferrucci. 2017. Gender Games: The portrayal of female journalists on House of Cards. *Journalism Practice*, 11(4):493–508.

Marcin Skowron, Martin Trapp, Sabine Payr, and Robert Trappl. 2016. Automatic Identification of Character Types from Film Dialogs. *Applied Artificial Intelligence*, 30(10):942–973.