Analyzing Gender Stereotyping in Bollywood Movies

Nishtha Madaan[1], Sameep Mehta[1], Taneea S Agrawaal[2], Vrinda Malhotra[2], Aditi Aggarwal[2], Mayank Saxena[3]

[1]IBM Research-INDIA

[2]IIIT-Delhi

[3]DTU-Delhi

{nishthamadaan, sameepmehta}@in.ibm.com , {taneea14166, vrinda14122, aditi16004}@iiitd.ac.in, {mayank26saxena}@gmail.com

Abstract

The presence of gender stereotypes in many aspects of society is a well-known phenomenon. In this paper, we focus on studying such stereotypes and bias in Hindi movie industry (*Bollywood*). We analyze movie plots and posters for all movies released since 1970. The gender bias is detected by semantic modeling of plots at inter-sentence and intra-sentence level. Different features like occupation, introduction of cast in text, associated actions and descriptions are captured to show the pervasiveness of gender bias and stereotype in movies. We derive a semantic graph and compute centrality of each character and observe similar bias there. We also show that such bias is not applicable for movie posters where females get equal importance even though their character has little or no impact on the movie plot. Furthermore, we explore the movie trailers to estimate on-screen time for males and females and also study the portrayal of emotions by gender in them. The silver lining is that our system was able to identify 30 movies over last 3 years where such stereotypes were broken.

Introduction

Movies are a reflection of the society. They mirror (with creative liberties) the problems, issues, thinking & perception of the contemporary society. Therefore, we believe movies could act as the proxy to understand how prevalent gender bias and stereotypes are in any society. In this paper, we leverage NLP and image understanding techniques to quantitatively study this bias. To further motivate the problem we pick a small section from the plot of a blockbuster movie.

"Rohit is an aspiring singer who works as a salesman in a car showroom, run by Malik (Dalip Tahil). One day he meets Sonia Saxena (Ameesha Patel), daughter of Mr. Saxena (Anupam Kher), when he goes to deliver a car to her home as her birthday present."

This piece of text is taken from the *plot* of Bollywood movie *Kaho Na Pyaar Hai*. This simple two line plot showcases the issue in following fashion:

1. Male (Rohit) is portrayed with a profession & an aspiration

2. Male (Malik) is a business owner

In contrast, the female role is introduced with no profession or aspiration. The introduction, itself, is dependent upon another male character *"daughter of"*!

One goal of our work is to analyze and quantify gender-based stereotypes by studying the demarcation of roles designated to males and females. We measure this by performing an intra-sentence and inter-sentence level analysis of movie plots combined with the cast information. Capturing information from sentences helps us perform a holistic study of the corpus. Also, it helps us in capturing the characteristics exhibited by male and female class. We have extracted movies pages of all the Hindi movies released from 1970-present from Wikipedia. We also employ deep image analytics to capture such bias in movie posters and previews.

Analysis Tasks

We focus on following tasks to study gender bias in Bollywood.

I) **Occupations and Gender Stereotypes**- How are males portrayed in their jobs vs females? How are these levels different? How does it correlate to gender bias and stereotype?

II) **Appearance and Description** - How are males and females described on the basis of their appearance? How do the descriptions differ in both of them? How does that indicate gender stereotyping?

III) **Centrality of Male and Female Characters** - What is the role of males and females in movie plots? How does the amount of male being central or female being central differ? How does it present a male or female bias?

IV) **Mentions(Image vs Plot)** - How many males and females are the faces of the promotional posters? How does this correlate to them being mentioned in the plot? What results are conveyed on the combined analysis?

V) **Dialogues** - How do the number of dialogues differ between a male cast and a female cast in official movie script?

VI) **Singers** - Does the same bias occur in movie songs? How does the distribution of singers with gender vary over a period of time for different movies?

VII) **Female-centric Movies**- Are the movie stories and portrayal of females evolving? Have we seen female-centric movies in the recent past?

VIII) **Screen Time** - Which gender, if any, has a greater screen time in movie trailers?

IX) **Emotions of Males and Females** - Which emotions are most commonly displayed by males and females in a movie trailer? Does this correspond with the gender stereotypes which exist in society?

Related Work

While there are recent works where gender bias has been studied in different walks of life (Soklaridis et al. 2017),(**?**), (Carnes et al. 2015), (Terrell et al. 2017), (Saji 2016), the analysis majorly involves information retrieval tasks involving a wide variety of prior work in this area. (Fast, Vachovsky, and Bernstein 2016) have worked on gender stereotypes in English fiction particularly on the Online Fiction Writing Community. The work deals primarily with the analysis of how males and females behave and are described in this online fiction. Furthermore, this work also presents that males are over-represented and finds that traditional gender stereotypes are common throughout every genre in the online fiction data used for analysis.

Apart from this, various works where Hollywood movies have been analyzed for having such gender bias present in them (Anderson and Daniels 2017). Similar analysis has been done on children books (Gooden and Gooden 2001) and music lyrics (Millar 2008) which found that men are portrayed as strong and violent, and on the other hand, women are associated with home and are considered to be gentle and less active compared to men. These studies have been very useful to uncover the trend but the derivation of these analyses has been done on very small data sets. In some works, gender drives the decision for being hired in corporate organizations (Dobbin and Jung 2012). Not just hiring, it has been shown that human resource professionals' decisions on whether an employee should get a raise have also been driven by gender stereotypes by putting down female claims of raise requests. While, when it comes to consideration of opinion, views of females are weighted less as compared to those of men (Otterbacher 2015). On social media and dating sites, women are judged by their appearance while men are judged mostly by how they behave (Rose et al. 2012; Otterbacher 2015; Fiore et al. 2008). When considering occupation, females are often designated lower level roles as compared to their male counterparts in image search results of occupations (Kay, Matuszek, and Munson 2015). In our work we extend these analyses for Bollywood movies.

The motivation for considering Bollywood movies is three fold:

a) The data is very diverse in nature. Hence finding how gender stereotypes exist in this data becomes an interesting study.

b) The data-set is large. We analyze 4000 movies which cover all the movies since 1970. So it becomes a good first step to develop computational tools to analyze the existence of stereotypes over a period of time.

c) These movies are a reflection of society. It is a good first step to look for such gender bias in this data so that necessary steps can be taken to remove these biases.

Data and Experimental Study

Data Selection

We deal with (three) different types of data for Bollywood Movies to perform the analysis tasks-

Movies Data   Our data-set consist of all Hindi movie pages from Wikipedia. The data-set contains 4000 movies for 1970-2017 time period. We extract movie title, cast information, plot, soundtrack information and images associated for each movie. For each listed cast member, we traverse their wiki pages to extract gender information. Cast Data consists of data for 5058 cast members who are Females and 9380 who are Males. *Since we did not have access to too many official scripts, we use Wikipedia plot as proxy. We strongly believe that the Wikipedia plot represent the correct story line. If an actor had an important role in the movie, it is highly unlikely that wiki plot will miss the actor altogether.*

Movies Scripts Data   We obtained PDF scripts of 13 Bollywood movies which are available online. The PDF scripts are converted into structured HTML using (Machines 2017). We use these HTML for our analysis tasks.

Movie Preview Data   Our data-set consists of 880 official movie trailers of movies released between 2008 and 2017. These trailers were obtained from YouTube. The mean and standard deviation of the duration of the all videos is 146 and 35 seconds respectively. The videos have a frame rate of 25 FPS and a resolution of 480p. Each 25th frame of the video is extracted and analyzed using face classification for gender and emotion detection (Octavio Arriaga 2017).

Task and Approach

In this section, we discuss the tasks we perform on the movie data extracted from Wikipedia and the scripts. Further, we define the approach we adopt to perform individual tasks and then study the inferences. At a broad level, we divide our analysis in four groups. These can be categorized as follows-

a) *At intra-sentence level* - We perform this analysis at a sentence level where each sentence is analyzed independently. We do not consider context in this analysis.

b) *At inter-sentence level* - We perform this analysis at a multi-sentence level where we carry context from a sentence to other and then analyze the complete information.

c) *Image and Plot Mentions* - We perform this analysis by correlating presence of genders in movie posters and in plot mentions.

d) *At Video level* - We perform this analysis by doing gender and emotion detection on the frames for each video. (Octavio Arriaga 2017)

We define different tasks corresponding to each level of analysis.

Tasks at Intra-Sentence level   To make plots analysis ready, we used OpenIE (Fader, Soderland, and Etzioni 2011) for performing co-reference resolution on movie plot text. The co-referenced plot is used for all analyses.

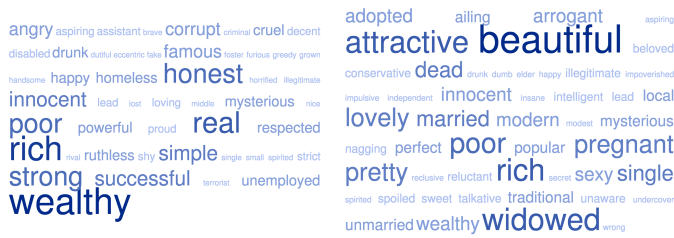The following intra-sentence analysis is performed

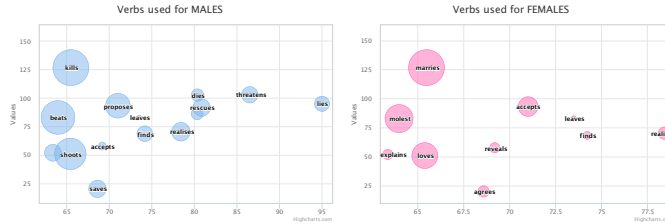Figure 1: Adjectives used with males and females

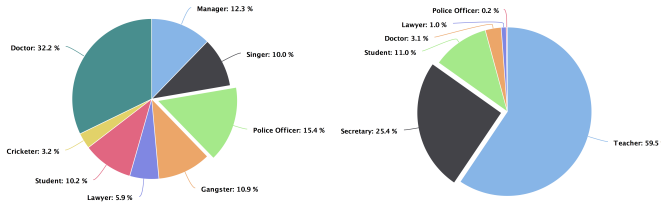Figure 2: Verbs used with males and females

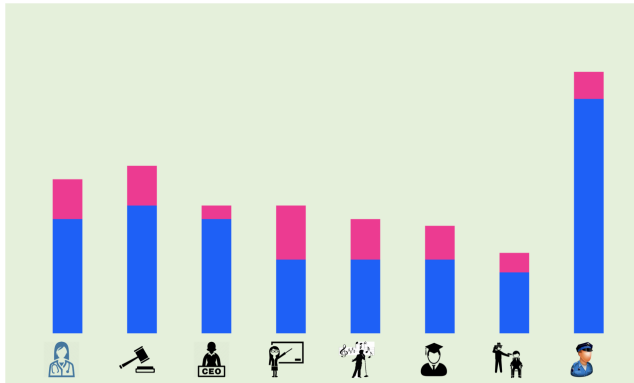Figure 3: Occupations of males and females

Figure 4: Gender-wise Occupations in Bollywood movies

1) **Cast Mentions in Movie Plot** - We extract mentions of male and female cast in the co-referred plot. The motivation to find mentions is how many times males have been referred to in the plot versus how many times females have been referred to in the plot. This helps us identify if the actress has an important role in the movie or not. In Figure 5 it is observed that, a male is mentioned around 30 times in a plot while a female is mentioned only around 15 times. *Moreover, there is a consistency of this ratio from 1970 to 2017(for almost 50 years)!*

2) **Cast Appearance in Movie Plot** - We analyze how male cast and female cast have been addressed. This essentially involves extracting verbs and adjectives associated with male cast and female cast. To extract verbs and adjectives linked to a particular cast, we use Stanford Dependency Parser (De Marneffe et al. 2006). In Fig 1 and 2 we present the adjectives and verbs associated with males and females. We observe that, verbs like *kills, shoots* occur with males while verbs like *marries, loves* are associated with females. Also when we look at adjectives, males are often represented as rich and wealthy while females are represented as beautiful and attractive in movie plots.

3) **Cast Introductions in Movie Plot** - We analyze how male cast and female cast have been introduced in the plot. We use OpenIE (Fader, Soderland, and Etzioni 2011) to capture such introductions by extracting relations corresponding to a cast. Finally, on aggregating the relations by gender, we find that males are generally introduced with a profession like as *a famous singer, an honest police officer, a successful scientist and so on* while females are either introduced using physical appearance like *beautiful, simple looking* or in relation to another (male) character (daughter, sister of). The results show that females are always associated with a successful male and are not portrayed as independent while males are portrayed to be successful.

4) **Occupation as a stereotype** - We perform a study on how occupations of males and females are represented. To perform this analysis, we collated an occupation list from multiple sources over the web comprising of 350 occupations. We then extracted an associated "noun" tag attached with cast member of the movie using Stanford Dependency Parser (De Marneffe et al. 2006) which is later matched to the available occupation list. In this way, we extract occupations for each cast member. We group these occupations for male and female cast members for all the collated movies. Figure 3 shows the occupation distribution of males and females. From the figure it is clearly evident that, males are given higher level occupations than females. Figure 4 presents a combined plot of percentages of male and female having the same occupation. This plot shows that when it comes to occupation like "teacher" or "student", females are high in number. But for "lawyer" and "doctor" the story is totally opposite.

5) **Singers and Gender distribution in Soundtracks** - We perform an analysis on how gender-wise distribution of singers has been varying over the years. To accomplish this, we make use of Soundtracks data present for each movie. This data contains information about songs and their corresponding singers. We extracted genders for each listed singer using their Wikipedia page and then aggregated the numbers of songs sung by males and females over the years. In Figure 7, we report the aforementioned distribution for recent years ranging from 2010-2017. We observe that the gender-gap is almost consistent over all these years.

Please note that currently this analysis only takes into account the presence or absence of female singer in a song. If one takes into account the actual part of the song sung, this
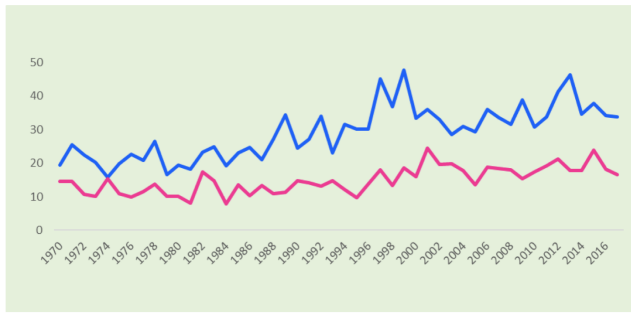
Figure 5: Total Cast Mentions showing mentions of male and female cast. Female mentions are presented in pink and Male mentions in blue

Figure 7: Gender-wise Distribution of singers in Sound-tracks
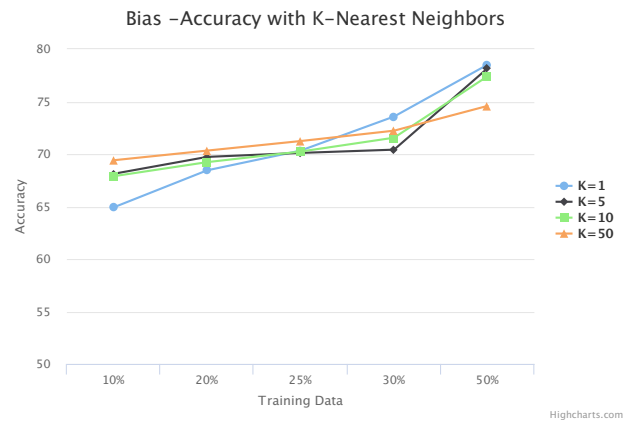
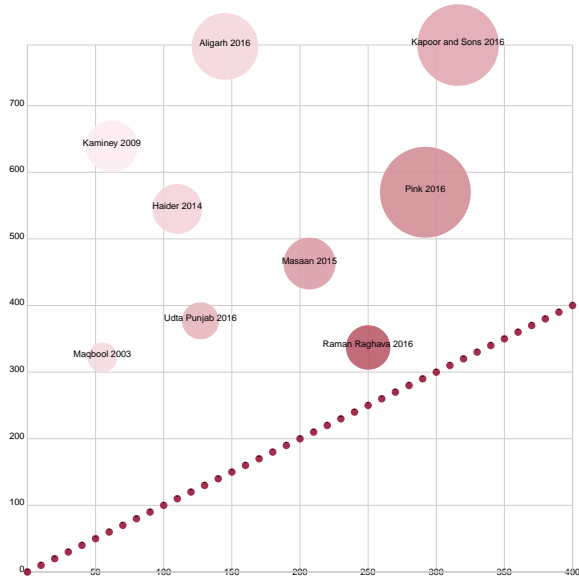Figure 8: Representing variation of Accuracy with training data

Figure 6: Total Cast dialogues showing ratio of male and female dialogues. Female dialogues are presented on X-axis and Male dialogues on Y-axis

trend will be more dismal. In fact, in a recent interview [1] this particular hypothesis is even supported by some of the top female singers in Bollywood. In future we plan to use audio based gender detection to further quantify this.

6) **Cast Dialogues and Gender Gap in Movie Scripts** - We perform a sentence level analysis on 13 movie scripts available online. We have worked with PDF scripts and extracted structured pieces of information using (Machines 2017) pipeline in the form of structured HTML. We further extract the dialogues for a corresponding cast and later group this information to derive our analysis.

We first study the ratio of male and female dialogues. In figure 6, we present a distribution of dialogues in males and females among different movies. X-Axis represents number
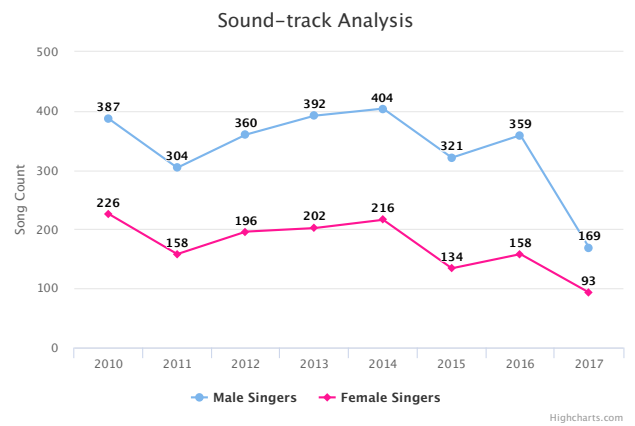
of female dialogues and Y-Axis represents number of male dialogues. The dotted straight line showing $y = x$. Farther a movie is from this line, more biased the movie is. In the figure 6, *Raman Raghav* exhibits least bias as the number of male dialogues and female dialogues distribution is not skewed. As opposed to this, *Kaminey* shows a lot of bias with minimal or no female dialogues.

Tasks at Inter-Sentence level We analyze the Wikipedia movie data by exploiting plot information. This information is collated at inter-sentence level to generate a context flow using a word graph technique. We construct a word graph for each sentence by treating each word in sentence as a node, and then draw grammatical dependencies extracted using Stanford Dependency Parser (De Marneffe et al. 2006) and connect the nodes in the word graph. Then using word graph for a sentence, we derive a knowledge graph for each cast member. The root node of knowledge graph is $[CastGender, CastName]$ and the relations represent the dependencies extracted using dependency parser across all sentences in the movie plot. This derivation is done by performing a merging step where we merge all the existing de-
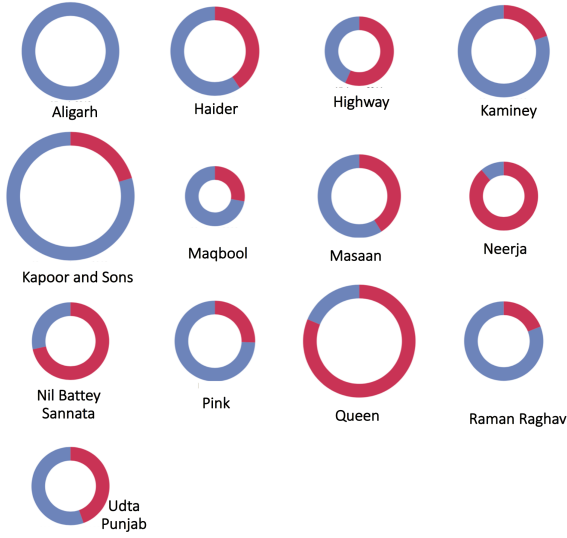
---

[1] goo.gl/BZWjWG

Figure 9: Lead Cast dialogues of males and females from different movie scripts

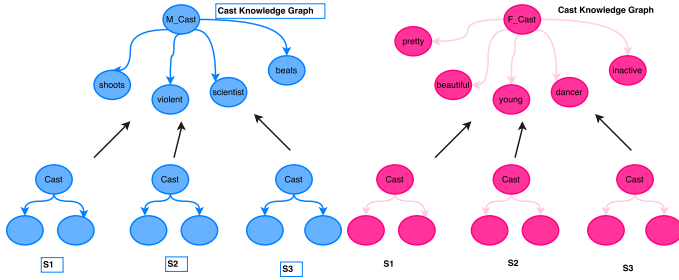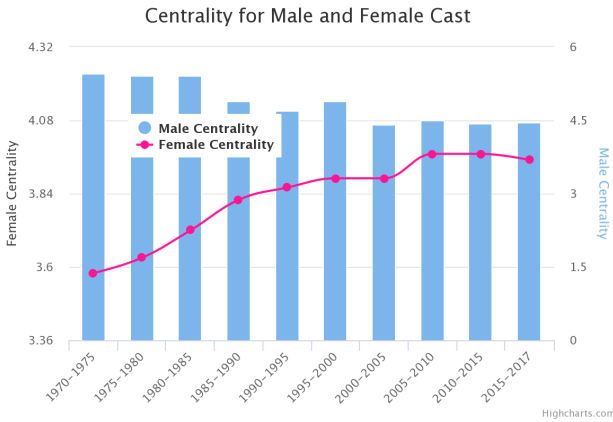

Figure 10: Knowledge graph for Male and Female Cast

Figure 11: Centrality for Male and Female Cast

pendencies of the cast node in all the word graphs of individual sentences. Figure 10 represents a sample knowledge graph constructed using individual dependencies.

After obtaining the knowledge graph, we perform the following analysis tasks on the data -

1. **Centrality of each cast node** - Centrality for a cast is a measure of how much the cast has been focused in the plot. For this task, we calculate between-ness centrality for cast node. Between-ness centrality for a node is number of shortest paths that pass through the node. We find between-ness centrality for male and female cast nodes and analyze the results. In Figure 11, we show male and female centrality trend across different movies over the years. We observe that there is a huge gap in centrality of male and female cast.

2. **Study of bias using word embeddings** - So far, we have looked at verbs, adjectives and relations separately. In this analysis, we want to perform joint modeling of aforementioned. For this analysis, we generated word vectors using Google word2vec (Mikolov et al. 2013) of length 200 trained on Bollywood Movie data scraped from Wikipedia. CBOW model is used for training Word2vec. The knowledge graph constructed for male and female cast for each movie contains a set of nodes connected to them. These nodes are extracted using dependency parser. We assign a context vector to each cast member node. The context vector consists of average of word vector of its connected nodes. As an instance, if we consider figure 10, the context vector for $[M, Castname]$ would be average of word vectors of (shoots, violent, scientist, beats). In this fashion we assign a context vector to each cast node.

The main idea behind assigning a context vector is to analyze the differences between contexts for male and female.

We randomly divide our data into training and testing data. We fit the training data using a K-Nearest Neighbor with varying K. We study the accuracy results by varying samples of train and test data. In Figure 8, we show the accuracy values for varying values of K. While studying bias using word embeddings by constructing a context vector, the key point is when training data is 10%, we get almost 65%-70% accuracy, refer to Figure 8. This pattern shows very high bias in our data. As we increase the training data, the accuracy also shoots up. There is a distinct demarcation in verbs, adjectives, relations associated with males and females. Although we did an individual analysis for each of the aforementioned intra-sentence level tasks, but the combined inter-sentence level analysis makes the argument of existence of bias stronger. Note the key point is not that the accuracy goes up as the training data is increased. The key point is that since the gender bias is high, the small training data has enough information to classify correctly 60-70% of the cases.

Movie Poster and Plot Mentions We analyze images on Wikipedia movie pages for presence of males and females on publicity posters for the movie. We use Dense CAP (Johnson, Karpathy, and Fei-Fei 2016) to extract male and female occurrences by checking our results in the top 5 responses having a positive confidence score.

After the male and female extraction from posters, we analyze the male and female mentions from the movie plot and co-relate them. The intent of this analysis is to learn how publicizing a movie is biased towards a female on advertising material like posters, and have a small or inconsequential role in the movie.
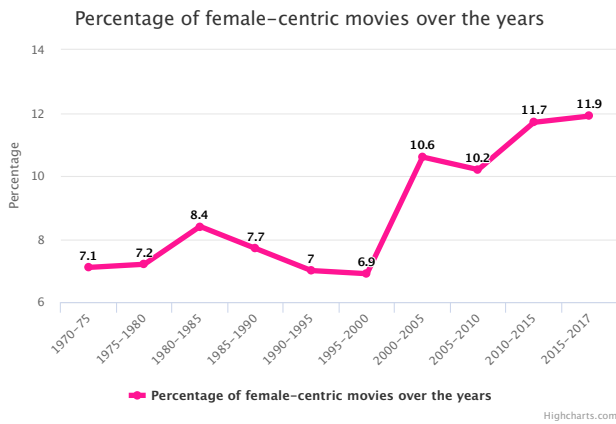
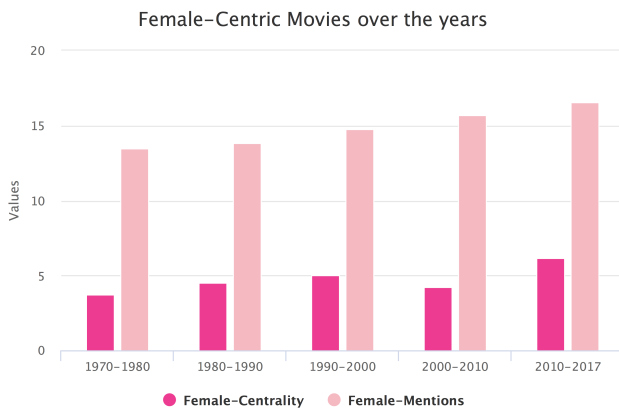Figure 12: Percentage of female-centric movies over the years

Figure 13: Percentage of female-centric movies over the years

While 80% of the movie plots have more male mentions than females, surprisingly more than 50% movie posters feature actresses. Movies like GangaaJal [2], Platform [3], Raees [4] have almost 100+ male mentions in plot but 0 female mentions whereas in all 3 posters females are shown on posters very prominently. Also, when we look at Image and Plot mentions, we observe that in 56% of the movies, female plot mentions are less than half the male plot mentions while in posters this number is around 30%. Our system detected 289 female-centric movies, where this stereotype is being broken. To further study this, we plotted centrality of females and their mentions in plots over the years for these 289 movies. Figure 13 shows that both plot mentions and female centrality in the plot exhibit an increasing trend which essentially means that there has been a considerable increase in female roles over the years. We also study the number of female-centric movies to the total movies over the years. Figure 12 shows the percentage chart and the trend for per-

---

[2]https://en.wikipedia.org/wiki/Gangaajal

[3]https://en.wikipedia.org/wiki/Platform_(1993_film)

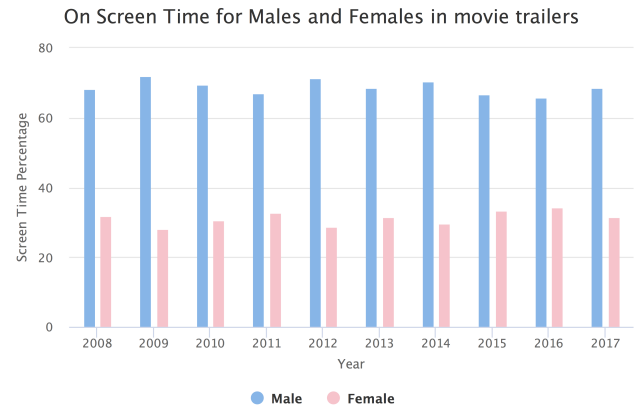[4]https://en.wikipedia.org/wiki/Raees_(film)

Figure 14: Percentage of screen-on time for males and females over the years

centage of female-centric movies. It is enlightening to see that the percentage shows a rising trend. Our system discovered at least 30 movies in last three years where females play central role in plot as well as in posters. We also note that over time such biases are decreasing - still far away from being neutral but the trend is encouraging. Figure 12 shows percentage of movies in each decade where women play more central role than male.

Movie Preview Analysis

We analyze all the frames extracted from the movie preview dataset and obtain information regarding the presence/absence of a male/female in the frame. If any person is present in the frame we then find out the emotion displayed by the person. The emotion displayed can be one of angry, disgust, fear, happy, neutral, sad, surprise. Note that there can be more than one person detected in a single frame, in that instance, emotions of each person is detected. We then aggregate the results to analyze the following tasks on the data -

1. Screen-On Time - Figure 14 shows the percentage distribution of screen-on time for males and female characters in movie trailers. We see a consistent trend across the 10 years where mean screen-on time for females is only a meagre 31.5 % compared to 68.5 % of the time for male characters.

2. Portrayal through Emotions - In this task we analyze the emotions most commonly exhibited by male and female characters in movie trailers. The most substantial difference is seen with respect to the "Anger" emotion. Over the 10 years, anger constitutes 26.3 % of the emotions displayed by male characters as compared to the 14.5 % of emotions displayed by female characters. Another trend which is observed, is that, female characters have always been shown as more happy than male characters every year. These results correspond to the gender stereotypes which exist in our society. We have not shown plots for other emotions because we could not see any proper trend exhibited by them.
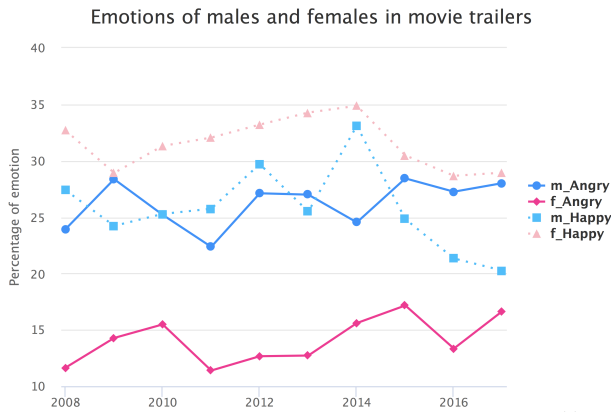
Figure 15: Year wise distribution of emotions displayed by males and females.

Dataset Release

We make the dataset accumulated for this analysis public.

1) Movies Dataset - We extracted data for 4000 movies with cast, soundtracks and plot text. Using OpenIE we generated co-referenced plots. This is located here [5]. Also, we extracted verbs, relations and adjectives corresponding to each cast and their gender. This information can be found here [6].

2) Movie Preview Dataset - We extracted 880 movie trailers from YouTube from 2008 and 2017. Further, we extracted frames from each trailer which can be found here [7]. Using this data we calculated the on-screen time and emotions. The detailed results can be found here [8].

Discussion and Ongoing Work

While our analysis points towards the presence of gender bias in Hindi movies, it is gratifying to see that the same analysis was able to discover the slow but steady change in gender stereotypes.

We would also like to point out that the goal of this study is not to criticize one particular domain. Gender bias is pervasive in all walks of life including but not limited to the Entertainment Industry, Technology Companies, Manufacturing Factories & Academia. In many cases, the bias is so deep rooted that it has become the norm. We truly believe that the majority of people displaying gender bias do it unconsciously. We hope that ours and more such studies will help people realize when such biases start to influence every day activities, communications & writings in an unconscious manner, and take corrective actions to rectify the same. Towards that goal, we are building a system which can

re-write stories in a gender neutral fashion. To start with we are focusing on two tasks:

a) **Removing Occupation Hierarchy** : It is common in movies, novel & pictorial depiction to show man as boss, doctor, pilot and women as secretary, nurse and stewardess. In this work, we presented occupation detection. We are extending this to understand hierarchy and then evaluate if changing genders makes sense or not. For example, while interchanging ({male, doctor}, {female, nurse}) to ({male, nurse}, {female, doctor}) makes sense but interchanging {male, gangster} to {female, gangster} may be a bit unrealistic.

b) **Removing Gender Bias from plots**: The question we are trying to answer is "If one interchanges all males and females, is the plot/story still possible or plausible?". For example, consider a line in plot "She gave birth to twins", of course changing this from she to he leads to impossibility. Similarly, there could be possible scenarios but may be implausible like the gangster example in previous paragraph. Solving these problems would require development of novel text algorithms, ontology construction, fact (possibility) checkers and implausibility checkers. We believe it presents a challenging research agenda while drawing attention to an important societal problem.

Conclusion

This paper presents an analysis study which aims to extract existing gender stereotypes and biases from Wikipedia Bollywood movie data containing 4000 movies. The analysis is performed at sentence at multi-sentence level and uses word embeddings by adding context vector and studying the bias in data. We observed that while analyzing occupations for males and females, higher level roles are designated to males while lower level roles are designated to females. A similar trend has been exhibited for *centrality* where females were less central in the plot vs their male counterparts. Also, while predicting gender using context word vectors, with very small training data, a very high accuracy is observed in gender prediction for test data reflecting a substantial amount of bias present in the data. We use this rich information extracted from Wikipedia movies to study the dynamics of the data and to further define new ways of removing such biases present in the data. As a part of future work, we aim to extract summaries from this data which are bias-free. In this way, the next generations would stop inheriting bias from previous generations. While the existence of gender bias and stereotype is experienced by viewers of Hindi movies, to the best of our knowledge this is first study to use computational tools to quantify and trend such biases.

---

[5]https://github.com/BollywoodMovieDataAAAI/Bollywood-Movie-Data/blob/master/wikipedia-data/coref_plot.csv

[6]https://github.com/BollywoodMovieDataAAAI/Bollywood-Movie-Data/blob/master/wikipedia-data/

[7]https://github.com/BollywoodMovieDataAAAI/Bollywood-Movie-Data/tree/master/trailer-data/individual-trailers-data

[8]https://github.com/BollywoodMovieDataAAAI/Bollywood-Movie-Data/tree/master/trailer-data

References

[Anderson and Daniels 2017] Anderson, H., and Daniels, M. 2017. https://pudding.cool/2017/03/film-dialogue/.

[Carnes et al. 2015] Carnes, M.; Devine, P. G.; Manwell, L. B.; Byars-Winston, A.; Fine, E.; Ford, C. E.; Forscher, P.; Isaac, C.; Kaatz, A.; Magua, W.; et al. 2015. Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Academic*

medicine: journal of the Association of American Medical Colleges 90(2):221.

[De Marneffe et al. 2006] De Marneffe, M.-C.; MacCartney, B.; Manning, C. D.; et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, 449–454. Genoa Italy.

[Dobbin and Jung 2012] Dobbin, F., and Jung, J. 2012. Corporate board gender diversity and stock performance: The competence gap or institutional investor bias?

[Fader, Soderland, and Etzioni 2011] Fader, A.; Soderland, S.; and Etzioni, O. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1535–1545. Association for Computational Linguistics.

[Fast, Vachovsky, and Bernstein 2016] Fast, E.; Vachovsky, T.; and Bernstein, M. S. 2016. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In *ICWSM*, 112–120.

[Fiore et al. 2008] Fiore, A. T.; Taylor, L. S.; Mendelsohn, G. A.; and Hearst, M. 2008. Assessing attractiveness in online dating profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 797–806. ACM.

[Gooden and Gooden 2001] Gooden, A. M., and Gooden, M. A. 2001. Gender representation in notable children's picture books: 1995–1999. *Sex roles* 45(1-2):89–101.

[Johnson, Karpathy, and Fei-Fei 2016] Johnson, J.; Karpathy, A.; and Fei-Fei, L. 2016. Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4565–4574.

[Kay, Matuszek, and Munson 2015] Kay, M.; Matuszek, C.; and Munson, S. A. 2015. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3819–3828. ACM.

[Machines 2017] Machines, I. B. 2017. https://www.ibm.com/watson/developercloud/developer-tools.html.

[Mikolov et al. 2013] Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[Millar 2008] Millar, B. 2008. Selective hearing: gender bias in the music preferences of young adults. *Psychology of music* 36(4):429–445.

[Octavio Arriaga 2017] Octavio Arriaga, P. G. P. 2017. Real-time convolutional neural networks for emotion and gender classification.

[Otterbacher 2015] Otterbacher, J. 2015. Linguistic bias in collaboratively produced biographies: crowdsourcing social stereotypes? In *ICWSM*, 298–307.

[Rose et al. 2012] Rose, J.; Mackey-Kallis, S.; Shyles, L.; Barry, K.; Biagini, D.; Hart, C.; and Jack, L. 2012. Face it: The impact of gender on social media images. *Communication Quarterly* 60(5):588–607.

[Saji 2016] Saji, T. 2016. Gender bias in corporate leadership: A comparison between indian and global firms. *Effective Executive* 19(4):27.

[Soklaridis et al. 2017] Soklaridis, S.; Kuper, A.; Whitehead, C.; Ferguson, G.; Taylor, V.; and Zahn, C. 2017. Gender bias in hospital leadership: a qualitative study on the experiences of women ceos. *Journal of Health Organization and Management* 31(2).

[Terrell et al. 2017] Terrell, J.; Kofink, A.; Middleton, J.; Rainear, C.; Murphy-Hill, E.; Parnin, C.; and Stallings, J. 2017. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science* 3:e111.