Proceedings of Machine Learning Research 81:1–14, 2018 Conference on Fairness, Accountability, and Transparency

Analyze, Detect and Remove Gender Stereotyping from Bollywood Movies

Nishtha Madaan NISHTHAMADAAN@IN.IBM.COM
Sameep Mehta SAMEEPMEHTA@IN.IBM.COM
IBM Research-INDIA

Taneea S Agrawaal TANEEA14166@IIITD.AC.IN
Vrinda Malhotra VRINDA14122@IIITD.AC.IN
Aditi Aggarwal ADITI16004@IIITD.AC.IN
IIIT-Delhi

Yatin Gupta YATINGUPTA12@GMAIL.COM
MSIT Delhi

Mayank Saxena MAYANK26SAXENA@GMAIL.COM
DTU-Delhi

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*) and propose an algorithm to remove these stereotypes from text. We analyze movie plots and posters for all movies released since 1970. The gender bias is detected by semantic modeling of plots at sentence and intra-sentence level. Different features like occupation, introductions, associated actions and descriptions are captured to show the pervasiveness of gender bias and stereotype in movies. Using the derived semantic graph, we 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. The silver lining is that our system was able to identify 30 movies over last 3 years where such stereotypes were broken. The next step, is to generate debiased stories. The proposed debiasing algorithm extracts gender biased graphs from unstructured piece of text in stories from movies and de-bias these graphs to generate plausible unbiased stories.

1. 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 demar-

© 2018 N. Madaan, S. Mehta, T.S. Agrawaal, V. Malhotra, A. Aggarwal, Y. Gupta & M. Saxena.

cation 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.

1.1. 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?

2. Related Work

While there are recent works where gender bias has been studied in different walks of life Soklaridis et al. (2017),(MacNell et al., 2015), (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 et al., 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 et al., 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.

3. Data and Experimental Study

3.1. Data Selection

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

3.1.1. 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.*

3.1.2. 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.

3.1.3. 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 25^{th} frame of the video is extracted and analyzed using face classification for gender and emotion detection (Octavio Arriaga, 2017).

3.2. 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.

3.2.1. Tasks at Intra-Sentence level

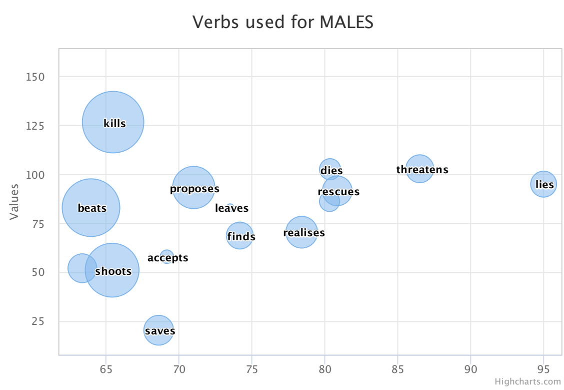
To make plots analysis ready, we used OpenIE (Fader et al., 2011) for performing co-reference resolution on movie plot text. The co-referenced plot is used for all analyses.

The following intra-sentence analysis is performed

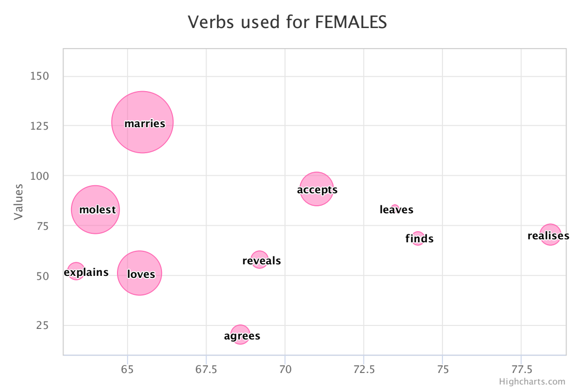
1) **Cast Mentions in Movie Plot** - We extract mentions of male and female cast in the co-referred plot. The motivation to find mentions is how many times males have been referred to in the plot versus how many times females have been referred to in the plot. This helps us identify if the actress has an important role in the movie or not. In Figure 2 it is observed that, a male is mentioned around 30 times in a plot

(*a*) Adjectives used with males

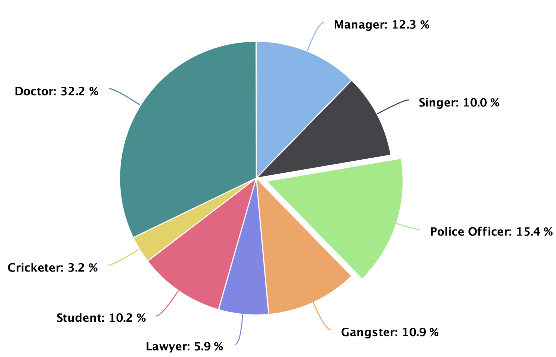
(*b*) Adjectives used with females



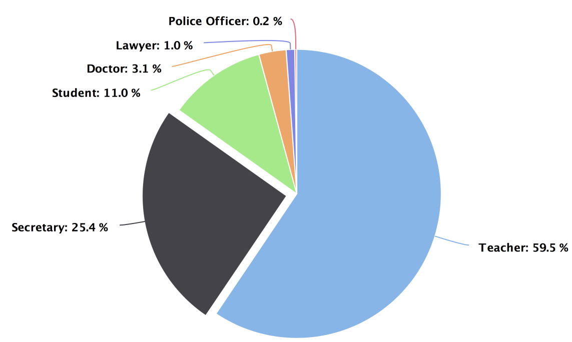
(*c*) Verbs used with males



(*d*) Verbs used with females



(*e*) Occupations used with males



(*f*) Occupations used with females

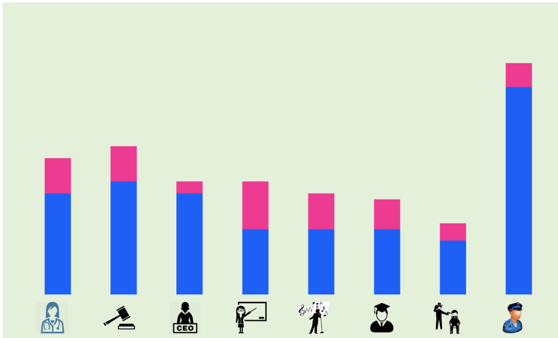


Figure 1: Gender-wise Occupations in Bollywood movies

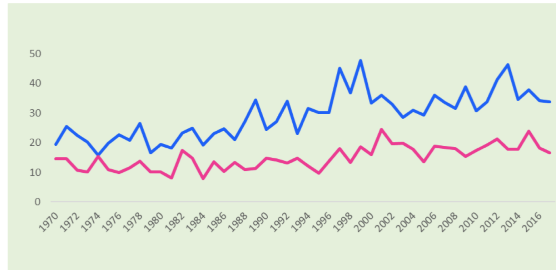


Figure 2: Total Cast Mentions showing mentions of male and female cast. Female mentions are presented in pink and Male mentions in blue

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 **??** and **??** 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 et al., 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 **??** 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 1 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 4, 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

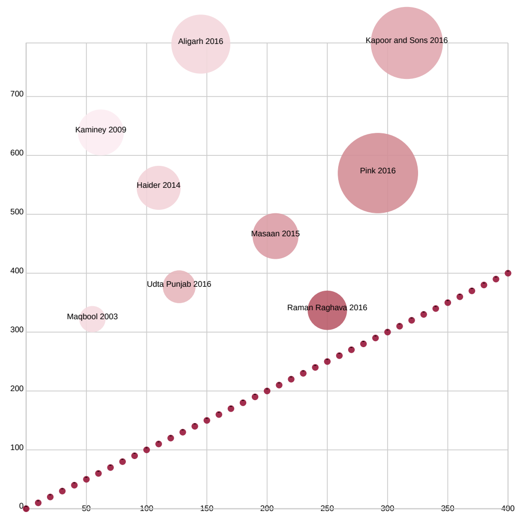


Figure 3: Total Cast dialogues showing ratio of male and female dialogues. Female dialogues are presented on X-axis and Male dialogues on Y-axis

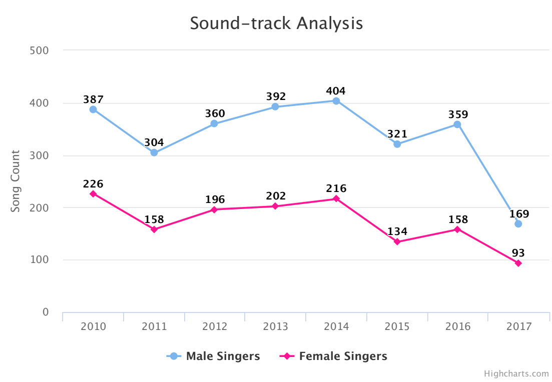


Figure 4: Gender-wise Distribution of singers in Soundtracks

actual part of the song sung, this 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.

1. goo.gl/BZWjWG

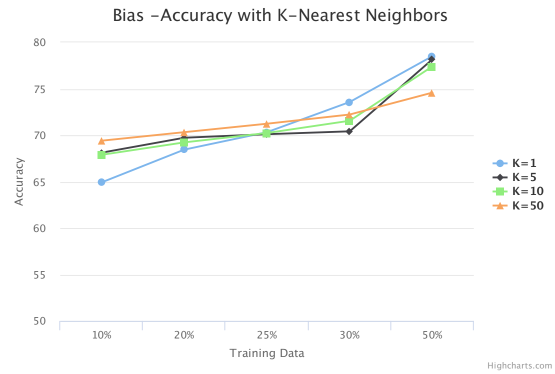


Figure 5: Representing variation of Accuracy with training data

We first study the ratio of male and female dialogues. In figure 3, 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 3, *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.

3.2.2. 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

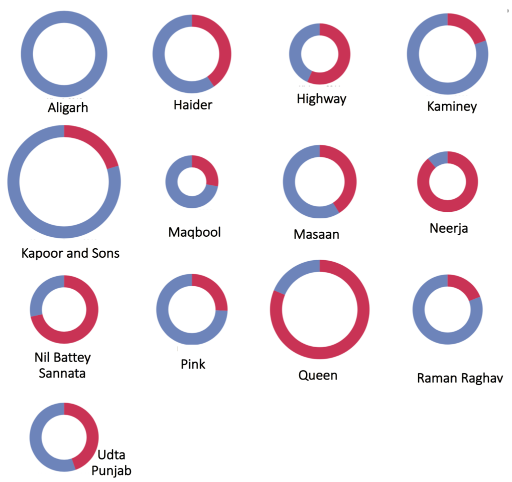
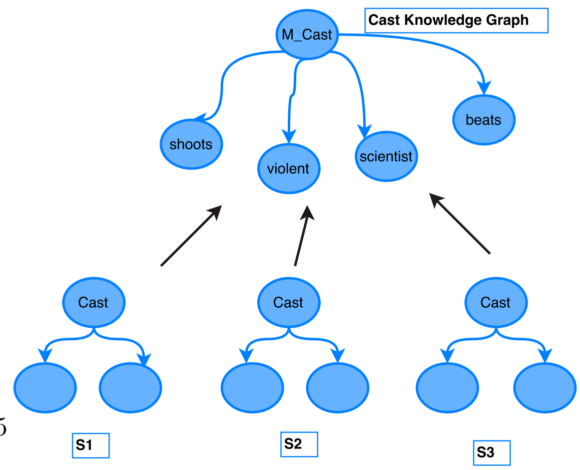


Figure 6: Lead Cast dialogues of males and females from different movie scripts



0.25
(a)

0.25
(b)

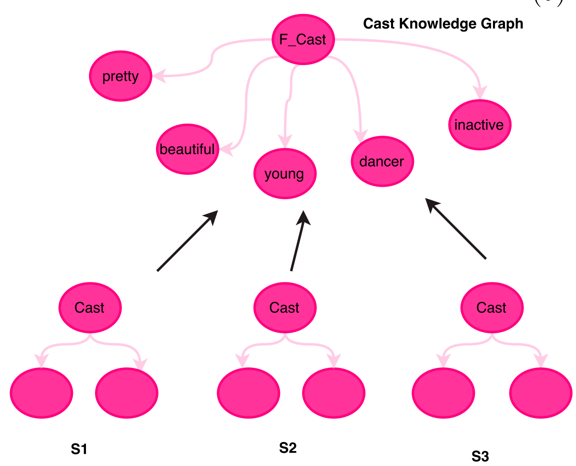


Figure 7: Knowledge graph for Male and Female Cast

merging step where we merge all the existing dependencies of the cast node in all the word graphs of individual sentences. Figure 7 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 8, 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 7, 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

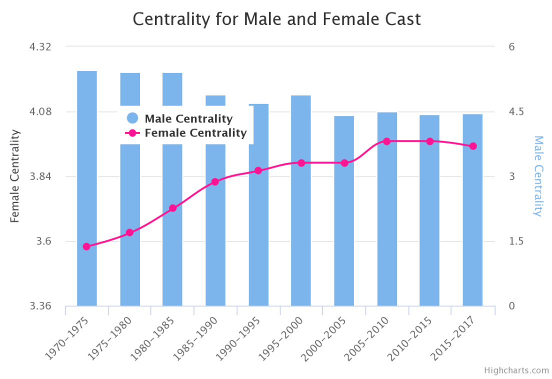


Figure 8: Centrality for Male and Female Cast

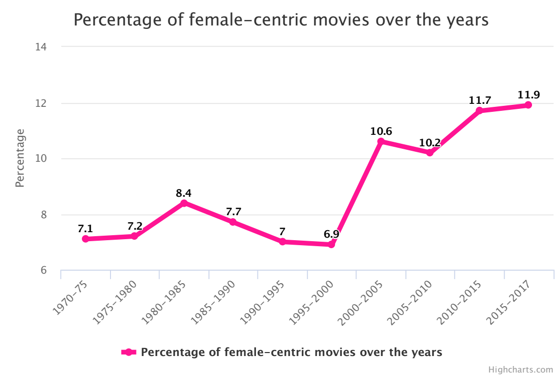


Figure 9: Percentage of female-centric movies over the years

test data. In Figure 5, 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 5. 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.

3.2.3. 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 et al., 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.

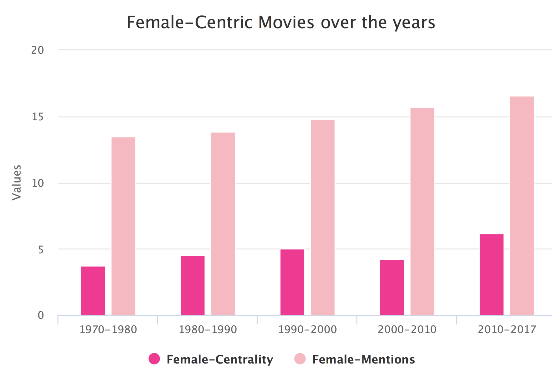


Figure 10: Percentage of female-centric movies over the years

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 10 shows that both plot mentions and female centrality in the plot exhibit

2. https://en.wikipedia.org/wiki/Gangaajal
3. https://en.wikipedia.org/wiki/Platform_(1993_film)
4. https://en.wikipedia.org/wiki/Raees_(film)

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 9 shows the percentage chart and the trend for percentage 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 9 shows percentage of movies in each decade where women play more central role than male.

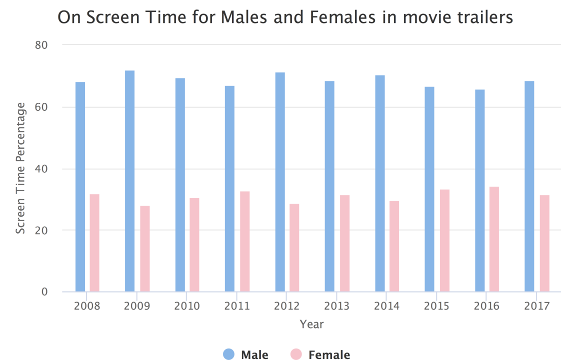


Figure 11: Percentage of screen-on time for males and females over the years

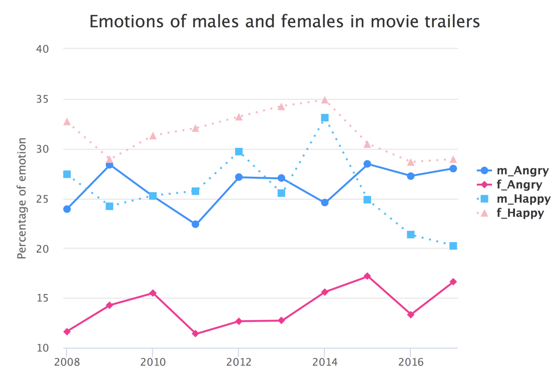


Figure 12: Year wise distribution of emotions displayed by males and females.

3.3. 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 11 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.

4. Algorithm for Bias Removal System - DeCogTeller

For this task, we take a news articles data set and train word embedding using Google *word2vec* (Mikolov et al., 2013). This data acts as a *fact data* which is used later to check for gender specificity of a particular action as per the facts. Apart from interchanging the actions, we have developed a specialized module to handle occupations. Very often, gender bias shows in assigned occupation { (Male, Doctor), (Female, Nurse)} or { (Male, Boss), (Female, Assistant)}.

In Figure 13 we give a holistic view of our system DeCogTeller which is described in a detailed manner as follows

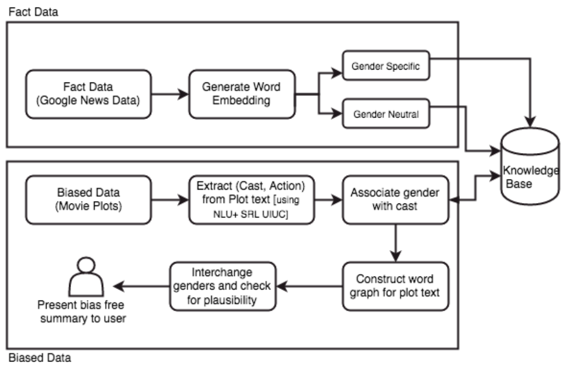


Figure 13: DeCogTeller- Bias Removal System

I) **Data Pre-processing** - We first perform data pre-processing of the words in fact data and do the following operations -
(a) used Wordnet to look-up if the word present in fact data is present in Wordnet(Miller, 1995) or not. If it was not present in Wordnet, the word was simply removed.
(b) used Stanford stemmer to stem the words so that the words like *modern, modernized etc.* don't form different vectors.

II) **Generating word vectors** - After we have the pre-processed list of words from fact data, we train Google word2vec and generate word embedding from this data. We do a similar operation on biased data which in our case is movies data from Bollywood.

III) **Extraction of analogical pairs** - The next task is to find analogical pairs from fact data which are analogous to the $(man, woman)$ pair. As an instance, if we take an analogical word pair (x, y) and we associate a vector $P(x, y)$ to the pair , then the task is to find
$P(x, y) = (vec[man] - vec[woman]) - (vec[x] - vec[y])$
Here, in the above equation we replace man and woman vectors by *he* and *she*, respectively. The above equation becomes
$P(x, y) = (vec[he] - vec[she]) - (vec[x] - vec[y])$
The main intent of this operation is to capture word pairs such as doctor or nurse where in most of the data, doctor is close to he and nurse is closer to she. Therefore for $(x, y) = (doctor, nurse)$,
$P(doctor, nurse)$ is given by $(vec[he]-vec[she])-(vec[doctor] - vec[nurse])$. Another example of (x, y) found in our data is $(king, queen)$. We generate all such (x, y) pairs and store them in our knowledge base. To have refined pairs, we used a scoring mechanism to filter important pairs. If

$$\|\mathbf{P}(x, y)\| \leq \tau$$

where τ is the threshold parameter, then add the word pair to knowledge base otherwise ignore. Equivalently, after normalizing $(vector[he] - vector[she])$ and $(vec[x] - vec[y])$, we calculated cosine distance as $cosine(vec[he] - vec[she], vec[x] - vec[y])$ which is algebraically equivalent to the above inequality.

IV) **Classifying word pairs** - After we identify analogical pairs, we observe that the degree of bias is still not known in each pair. So, we need to classify word pairs as specific to a gender or neutral to the gender. For example, Consider a word pair *(doctor, nurse)*, we know that whether male or female anyone can be a doctor or a nurse. Hence we call such a pair as gender neutral. On the contrary, if we consider a word pair *(king, queen)*, we know that king is associated with a male while queen is associated from a female. We call such word pairs as gender specific. Now, the task is to first find out which pairs extracted in the above step correspond to gender neutral and which ones correspond to gender specific. To do this, we first extract the words from knowledge base extracted from biased data and find how close they are to different genders. For a word w, we calculate cosine score of w with *he* as $cos(w, he)$. If w is very close to *he*, then it is specific to a man. Similarly for a word w', we do the similar operation for *she*. And if w' is very close to *she*, then it is specific to a woman. If a word w'' is almost equidistant from *he* and *she*, then it is labelled as gender neutral.

V) **Action Extraction from Biased Movie Data** - After we have gender specific and gender neutral words from the fact data, we work on the biased data to extract actions associated with movie cast. We extract gender for movie cast by crawling the corresponding Wikipedia pages for actors and actresses. After we have the corresponding gender for each cast in the movie, we perform co-referencing on the movie plot using *Stanford OpenIE* (Fader et al., 2011). Next, we collate actions corresponding to each cast using *IBM NLU API* (Machines, 2017) and *Semantic Role Labeler by UIUC* (Punyakanok et al., 2008).

VI) **Bias detection using Actions** - At this point we have the actions extracted from biased data corresponding to each gender. We can now use this data against fact data to check for bias. We will describe in the following system walk-through section how we use it on-the-fly to check for bias.

VII) **Bias Removal** - We construct a knowledge graph for each cast using relations from *Stanford dependency parser*. We use this graph to calculate the between-ness centrality for each cast and store these centrality scores in a knowledge base. We use the between-ness centrality score to interchange genders after we detect the bias.

5. Walk-through using an example

The system DeCogTeller takes in a text input from the user. The user starts entering a biased movie plot text for a movie, say, "Kaho na Pyar Hai" in Figure 14. This natural language text is submitted into the system in which, first, the text is co-referenced using *OpenIE*. Then, using *IBM NLU API* and *UIUC Semantic Role Labeller* actions pertaining to each cast are extracted and these are checked with gender specific and gender neutral lists. If for a corresponding cast_gender,action pair the corresponding vector is located in gender specific list then it can not be termed as a biased action. But on the other hand if a cast_gender,action pair occurring in the plot is not found in gender-specific but the opposite gender is found in gender-neutral list, then we tag the statement as a biased statement.

As an example text if the user enters - "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." At the very fist step, co-referencing is done which coverts the above text to - "Rohit is an aspiring singer who works as a salesman in a car showroom, run by Malik (Dalip Tahil). One day Rohit meets Sonia Saxena (Ameesha Patel), daughter of Mr. Saxena (Anupam Kher), when Rohit goes to deliver a car to her home as her birthday present." After this step, we extract actions corresponding to each cast and then check for bias. Here corresponding to cast Rohit we have the following actions - {*singer, salesman, meets, deliver*}. The gender for Rohit is detected by using wiki page of Hritik Roshan and is labelled as "male". We find actions corresponding to cast Sonia and find the following actions- {*daughter-of*}. Then we run our gender-specific and gender neutral checks and find that the actions are gender neutral. Hence there is a bias that exists. We do the similar thing for other cast members. Then, at the background, we extract highest centrality male and highest centrality female. And then switch their gender to generate de-biased plot. Figure 15 shows the de-biased plot. Also, there is an option given to the user to view the knowledge graphs for biased text and unbiased text to see how nodes in knowledge graph change.

6. 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,

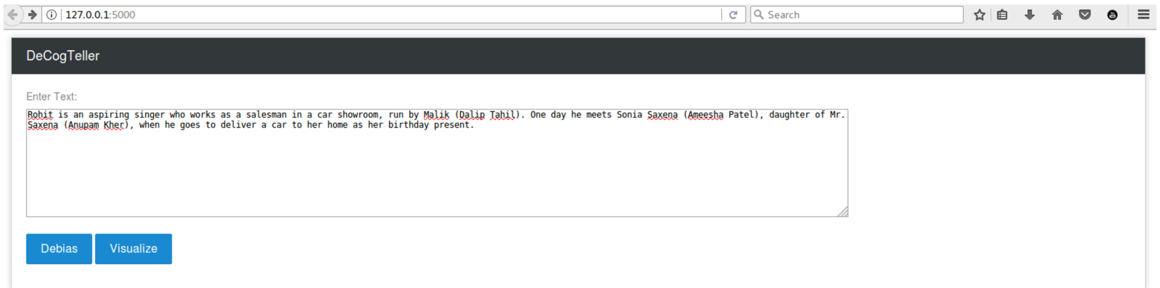
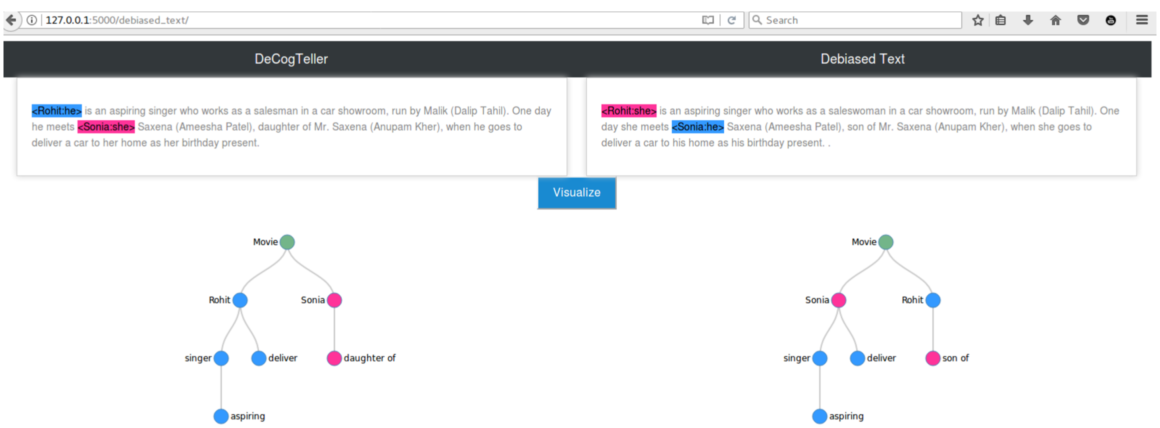


Figure 14: The screen where a user can enter the text



0.9
(a)

Figure 15: The screen where text is debiased and the knowledge graph can be visualized

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.

7. 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.

Furthermore, we present an algorithm to remove such bias present in text. We show that by interchanging the gender of high centrality male character with a high centrality female character in the plot text leaves no change in the story but de-biases it completely.

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.

References

Hanah Anderson and Matt Daniels. https://pudding.cool/2017/03/film-dialogue/. 2017.

Molly Carnes, Patricia G Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E Ford, Patrick Forscher, Carol Isaac, Anna Kaatz, Wairimu Magua, et al. 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, 2015.

Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454. Genoa Italy, 2006.

Frank Dobbin and Jiwook Jung. Corporate board gender diversity and stock performance: The competence gap or institutional investor bias? 2012.

Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics, 2011.

Ethan Fast, Tina Vachovsky, and Michael S Bernstein. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In *ICWSM*, pages 112–120, 2016.

Andrew T Fiore, Lindsay Shaw Taylor, Gerald A Mendelsohn, and Marti Hearst. Assessing attractiveness in online dating profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 797–806. ACM, 2008.

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

Justin Johnson, Andrej Karpathy, and Li Fei-Fei. Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4565–4574, 2016.

Matthew Kay, Cynthia Matuszek, and Sean A Munson. 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*, pages 3819–3828. ACM, 2015.

International Business Machines. https://www.ibm.com/watson/developercloud/developer-tools.html. 2017.

Lillian MacNell, Adam Driscoll, and Andrea N Hunt. What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291, 2015.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

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

George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11): 39–41, 1995.

Paul G. Ploger Octavio Arriaga. Real-time convolutional neural networks for emotion and gender classification. 2017.

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

Vasin Punyakanok, Dan Roth, and Wen-tau Yih. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2):257–287, 2008.

Jessica Rose, Susan Mackey-Kallis, Len Shyles, Kelly Barry, Danielle Biagini, Colleen Hart, and Lauren Jack. Face it: The impact of gender on social media images. *Communication Quarterly*, 60(5):588–607, 2012.

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

Sophie Soklaridis, Ayelet Kuper, Cynthia Whitehead, Genevieve Ferguson, Valerie Taylor, and Catherine Zahn. Gender bias in hospital leadership: a qualitative study on the experiences of women ceos. *Journal of Health Organization and Management*, 31(2), 2017.

Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science*, 3:e111, 2017.