

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
|  | | Analyze, Detect and Remove Gender Stereotyping fromBollywood Movies | | | | |  | |
|  |  | | | | | | |  |
|  | | | |  |  | | | |
|  | | | | Prince Nahar |  | | | |
|  | | | | 5 Jan 2023—ML Project— |  | | | |
|  | | |  | | |  | | |

The presence of gender stereotypes in many aspects of society is a well-known phenomenon. 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. 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 | | | | |  |

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

1. 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?
2. 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?
3. 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?
4. 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?
5. 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?
6. Female-centric Movies- Are the movie stories and portrayal of females evolving? Have we seen female-centric movies in the recent past?

2. **Data and Experimental Study:**

The dataset has Scripts of the movies, Trailers of the movies, Wikipedia data about the movies and Images in the movies and is available at:<https://github.com/BollywoodData/Bollywood-Data>.

3**. Cast Mentions in Movie 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. it is observed that, a male is mentioned around 30 times in a plot while a female is mentioned only around 15 times.

4. **Cast Appearance in Movie Plot:**

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.

5. **Occupation as a stereotype:**

males are given higher level occupations than females. It 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.

6**. Singers and Gender distribution in Soundtracks:**

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

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.

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 the knowledge graph constructed for male and female cast for each movie contains a set of nodes connected to them.

**3**. **Movie Poster and Plot Mentions:**

While 80% of the movie plots have more male mentions than females, surprisingly more than 50% movie posters feature actresses. Movies like GangaaJal, Platform, Raees 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%

**Jupiter Notebook walkthrough**

**steps**

1. **Importing all the relevant libraries**
2. **Import data from database**
3. **Analytic of the distribution of male / female emotion types..**
4. **Plotting graph of man and woman emotions** (angry, sad, ect)**.**
5. **Plotting graph of man and woman emotions over years**
6. **How are males and females described** (by adjective in plots ).
7. **Finding the top 5 description for male and female**
8. **Centrality of Male and Female Characters**
9. **Finding and comparing gender Centrality by percentage**

**10. Does the same bias occur in movie songs?** (NO)

**11. Finding and comparing gender bias in songs by percentage**

**12.** **the proposed machine learning solution**

**1.Applying SVM (liner) on song data for bias**

**2. Applying SVM (radical) on song data for bias**

**3. Applying SVM (liner) on Centrality data for bias**

**4. Applying SVM (radical) on Centrality data for bias**

**5.** **Random Forest Classifier on singers data .**

**Creating an function for user input to find bias or Not.**