Stereotype Analysis Project Report

Project Repository: Stereotype Analysis

Results are shown on github: https://github.com/sm22mtech12004/stereotypeanalysis.git

1. Introduction

Gender stereotypes in media, especially in Bollywood movies, have long been a subject of study. This project focuses on analyzing gender roles in Bollywood movies using data-driven approaches. The aim is to quantify gender dominance in movie plots, trailers, and posters, and propose solutions to reduce gender biases in these films.

2. Dataset Overview

The datasets used in this project include:

- **Movie Scripts:** Textual data representing the narrative and dialogues.
- **Movie Trailers:** Visual and audio data used for promoting the movies.
- **Wikipedia Data:** Structured and unstructured data from Wikipedia pages about the movies.
- **Movie Posters:** Visual data in the form of images used for marketing the movies.

3. Methodology

3.1 Centrality analysis:

1. **Data Loading:**
    - Loaded the `male_centrality.csv` and `female_centrality.csv` files into pandas DataFrames.
2. **Data Cleaning:**
    - Trimmed leading and trailing spaces from column names to ensure accurate processing.

Data Analysis

1. **Centrality Aggregation:**

    - Grouped the data by `Movie Name` and aggregated `Total Centrality` and `Average Centrality` for both male and female characters.
2. **Data Merging:**
    - Merged the aggregated centrality data for male and female characters on `Movie Name`.
3. **Gender Dominance Determination:**
    - Compared `Male Total Centrality` with `Female Total Centrality` for each movie to determine which gender was more dominant.

3.2 Gender Detection from Movie Posters

1. **Face Detection:**
    - Used the MTCNN algorithm to detect faces in movie posters.
2. **Gender Classification:**
    - Applied the VGG16 model, pre-trained on ImageNet, to classify detected faces as male or female.
3. **Bounding Box Analysis:**
    - Calculated the area of bounding boxes for male and female faces to determine visual gender dominance in movie posters.

3.3 Sentiment Analysis

1. **Sentiment Scoring:**
    - Analyzed the sentiment of adverbs associated with female characters from 1970 to 2017 using the VADER sentiment analyzer.
    - Grouped the sentiment data by year and visualized the trends over time.

4. Results

4.1 Gender Dominance in Centrality Metrics

- The analysis showed varying degrees of male and female dominance across different movies. Some movies had a higher `Male Total Centrality`, indicating a male-dominant narrative, while others were female-dominant.

4.2 Gender Dominance in Movie Posters

- By analyzing the sizes of bounding boxes for male and female faces in movie posters, it was observed that some posters were visually dominated by male characters, while others featured more female characters

.

4.3 Sentiment Analysis of Adverbs

- The sentiment analysis of adverbs used in descriptions of female characters revealed trends in how female characters were portrayed over time. This analysis helps in understanding the emotional tone and stereotypes associated with female characters in Bollywood movies.
- Also we have female and Male song count year wise. If we get an analysis, we found that, No. of female song counts was higher than male song counts before 1990.
- From sentiment analysis year wise, we have results which states , Female positive sentiment score was higher than male positive sentiment score before 1990.

Pre-1990: Female characters generally had higher positive sentiment compared to male characters, suggesting that they were often portrayed in a more positive light.

Post-1990 Shift: There's a noticeable shift post-1990, where male characters began to dominate in positive sentiment, while female characters showed a rise in negative sentiment. Additionally, female characters started to exhibit higher neutral sentiment, indicating a move towards more balanced or less emotionally charged portrayals.

Recent Trends: In recent decades, male characters have maintained or increased their positive and neutral sentiment scores, while female characters have experienced a decline in positive sentiment and a rise in negative sentiment.

-

6. Implication :

This project provides a comprehensive analysis of gender stereotypes in Bollywood movies using various machine learning techniques. The results highlight the presence of gender biases in both the narrative structure and visual representation of movies. The proposed recommendations aim to foster a more inclusive and balanced portrayal of genders in Bollywood films

These plots visually demonstrate a shift in the portrayal of male and female characters in Bollywood over time. The declining positive sentiment and rising negative sentiment for female characters, coupled with the increase in male positive sentiment, reflect a potential increase in gender bias in storytelling. These insights could be crucial for content creators and filmmakers who are striving to create more balanced and equitable character portrayals in cinema.

By addressing these trends, filmmakers can work towards narratives that equally empower both male and female characters, ensuring that positive representation is maintained across all genders.

The analysis of song counts by gender over the years provides significant insights into the shifting trends in Bollywood's portrayal and emphasis on male and female characters through music:

1. **Dominance of Female Songs Pre-1990:**

Increased Focus on Female Narratives: The higher number of female songs before 1990 suggests that Bollywood movies of that era often featured female-centric narratives or placed significant emphasis on female voices. This could indicate a cultural or industry preference for highlighting female experiences and emotions through music.

2. **Shift Post-1990:**

Decline in Female Representation: After 1990, there was a noticeable decline in the number of songs sung by female artists compared to male artists. This shift may reflect a broader trend in Bollywood where male characters began to take center stage, both in narrative and musical representation.

The song count analysis reveals a significant shift in gender representation in Bollywood music, with a decline in female songs post-1990. This trend has implications for the overall representation of female characters in Bollywood films, suggesting a need for the industry to address this imbalance. By re-emphasizing female voices in movie soundtracks, Bollywood can work towards more equitable and inclusive narratives that reflect the diverse experiences of all characters.

The analysis of gender detection and recognition in movie posters reveals significant implications for how male and female characters are visually represented in Bollywood cinema. The prominence of male characters and the marginalization of female characters on posters can reinforce gender stereotypes and influence audience perceptions.

.

7. Future Work

- **Enhanced Models:**

  Improve the accuracy of gender classification and sentiment analysis models.

- **Using NLP solution as LLM:**

  In the analysis of gender stereotypes within Bollywood movie scripts, the use of advanced Natural Language Processing (NLP) techniques is essential. Large Language Models (LLMs) offer sophisticated capabilities to summarize text, understand context, and detect subtle biases in language. These models can provide deep insights into how male and female characters are portrayed, helping to identify and quantify gender-based stereotypes.

However, LLMs require significant computational resources and time, which can pose challenges, especially when dealing with large datasets like entire movie scripts. Despite these challenges, by adopting a strategic approach, we can effectively utilize LLMs while optimizing resource usage and ensuring timely analysis.

EXAMPLE CASE:

Based on the plot of the movie "Highway" as provided in the document, here is a summary:

Summary of the Movie "Highway":

"Highway" is a 2014 Bollywood film directed by Imtiaz Ali, starring Alia Bhatt and Randeep Hooda. The film follows the story of Veera Tripathi, a young woman from a wealthy family, who is kidnapped by a group of men led by Mahabir Bhati. Despite the initial fear and trauma, Veera finds a sense of freedom and solace in her captivity, far from the suffocating constraints of her privileged life.

The movie explores themes of freedom, self-discovery, and the complexity of human relationships. As Veera and Mahabir traverse the highways of North India, their journey takes them through various landscapes, reflecting the inner transformation both characters undergo. Veera, who initially tries to escape, gradually begins to connect with Mahabir and reveals the emotional scars from her past, including the trauma of childhood abuse.

Mahabir, who initially sees Veera as a hostage, begins to empathize with her and becomes protective of her. Their relationship evolves from captor and captive to one of mutual understanding and unspoken connection. The film culminates in a tragic yet liberating conclusion, as Veera ultimately gains the freedom she longed for, albeit at a great cost.

"Highway" is a poignant exploration of the human spirit, touching upon the themes of liberation from societal norms, the impact of past trauma, and the unexpected bonds that can form between two seemingly opposite individuals.

Gender Bias Analysis:

1. ***Female Character Development:***
    - ***Strong Female Protagonist:*** *Veera is portrayed as a complex and evolving character. She starts as a victim of both her captor and her past but grows into a more self-aware and liberated individual. This is a positive representation of a female character who undergoes significant personal growth and finds her own voice.*
    - ***Impact of Trauma:*** *The narrative acknowledges the impact of childhood trauma on Veera, highlighting a serious issue that affects many women. This portrayal is sensitive and aligns with realistic character development, though it centers on suffering as a key aspect of her identity.*
2. ***Male Character Development:***
    - ***Protective Male Role:*** *Mahabir's role evolves from being Veera's captor to her protector, showing a shift from a negative to a more positive portrayal. However, this dynamic could be seen as reinforcing a traditional gender role where the male character ultimately takes on a protective role.*

Sentiment Analysis:

1. ***Overall Sentiment:***
    - ***Mixed Sentiments:*** *The summary reflects a mix of positive, neutral, and negative sentiments. The positive sentiment arises from Veera's journey of self-discovery and liberation, while the negative sentiment stems from the trauma and harsh circumstances she faces, including her kidnapping and childhood abuse.*
    - ***Neutral Sentiments:*** *The neutral sentiments are seen in the narrative's focus on the complex, evolving relationship between Veera and Mahabir, which is neither wholly positive nor negative but rather deeply nuanced.*
2. ***Sentiment Breakdown:***
    - ***Positive Sentiment:***
        - *Words like "freedom," "self-discovery," "liberation," and "understanding" contribute to the positive sentiment, highlighting Veera's growth and the emotional depth of the film.*
    - ***Negative Sentiment:***
        - *Words like "kidnapped," "fear," "trauma," "abuse," and "tragic" contribute to the negative sentiment, reflecting the darker elements of Veera's journey and the harsh realities she faces.*

The sentiment analysis of the summary reflects a complex emotional landscape with a mix of positive, neutral, and negative sentiments. While Veera's character is portrayed positively in terms of growth and liberation, there are elements of traditional gender roles that could be seen as biased. The narrative highlights the dependence of female character development on male interactions and the portrayal of women primarily through the lens of suffering.

These insights could be valuable for filmmakers and writers in understanding how to create more balanced and nuanced representations of female characters, ensuring that their growth is depicted independently and not solely through their relationships with male characters

8. References

- MTCNN Face Detection
- VGG16 Model
- TextBlob for Sentiment Analysis
- VADER Sentiment Analysis
