

An Unfair Affinity Toward Fairness: Characterizing 70 Years of Social Biases in B^Hollywood (Student Abstract)

Kunal Khadilkar¹ and Ashiqur R. KhudaBukhsh¹

¹ Carnegie Mellon University
kkhadilk@cs.cmu.edu, akhudabu@cs.cmu.edu

Abstract

Bollywood, aka the Mumbai film industry, is one of the biggest movie industries in the world with a current movie market share of worth 2.1 billion dollars and a target audience base of 1.2 billion people. While the entertainment impact in terms of lives that Bollywood can potentially touch is mammoth, no NLP study on social biases in Bollywood content exists. We thus seek to understand social biases in a developing country through the lens of popular movies. Our argument is simple – popular movie content reflects social norms and beliefs in some form or shape. We present our preliminary findings on a longitudinal corpus of English subtitles of popular Bollywood movies focusing on (1) social bias toward a fair skin color (2) gender biases, and (3) gender representation. We contrast our findings with a similar corpus of Hollywood movies. Surprisingly, we observe that much of the biases we report in our preliminary experiments on the Bollywood corpus, also gets reflected in the Hollywood corpus.

Introduction

On a diachronic corpus of popular entertainment, what types of social biases can we analyze and detect? In this paper, we analyze a curated corpus of film subtitles from the Bollywood and Hollywood film industry, spread over 70 years and present our preliminary findings on three aspects: (1) social bias toward a fair skin color; (2) gender biases; and (3) gender representation. While studies analyzing gender stereotypes across different languages (Lewis and Lupyan 2020) and detecting bias in word embeddings (Garg et al. 2018) exist, Bollywood in NLP literature has received little focus (Khan and Taylor 2018; Madaan et al. 2018) and no existing work has considered the rich data set of subtitles contrasting the findings with a Hollywood subtitles corpus.

In Figure 1, what do the prominent presence of negative adjectives such as *wanton* in old Bollywood movies (see, Figure 1(a)) juxtaposed with the presence of positive verbs such as *respect* (see, Figure 1(b)) tell us? In this paper, we explore a recent technique (Palakodety, KhudaBukhsh, and Carbonell 2020) used to mine political insights in a novel context of uncovering social biases. Through a series of cloze tests (Taylor 1953) on a language model (Devlin

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

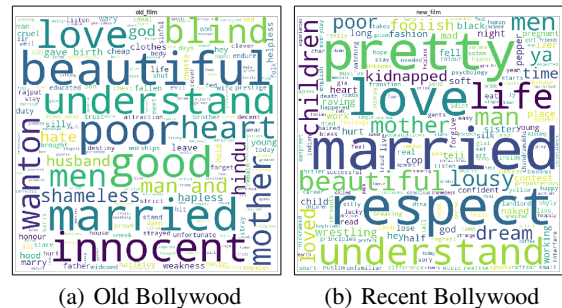


Figure 1: Word cloud visualizations for the five word contexts containing utterances of ‘woman/women’.

et al. 2019) fine-tuned on our data sets, we present our preliminary findings.

Our data sets: We construct the following two data sets.

1. Bollywood movies, $\mathcal{D}_{bollywood}$: We consider movies spanning seven decades (1950–2020). For each decade, we retrieved subtitles (Lison and Tiedemann 2016) of 40 top-grossing movies (280 total films), based on earnings documented by IMDb (Internet Movie Database).

2. Hollywood movies, $\mathcal{D}_{hollywood}$: Similarly, we considered 40 top-grossing movies from each of the seven decades (280 total films). Overall, $\mathcal{D}_{bollywood}$ and $\mathcal{D}_{hollywood}$ consist of 542,768 and 488,498 dialogues, respectively.

For a subset of our analyses, we divide our corpus into two buckets: (1) Films from 1950 to 1970 ($\mathcal{D}_{bollywood}^{old}$ and $\mathcal{D}_{hollywood}^{old}$); and (2) Films from 2000 to 2020 ($\mathcal{D}_{bollywood}^{new}$ and $\mathcal{D}_{hollywood}^{new}$).

Results

Pronouns as a Proxy for Representation: Following extensive literature on gendered pronouns’ relative distributions and their implications (Twenge, Campbell, and Gentile 2012), we consider a simple measure of gender representation: relative occurrence of pronouns of each gender (Men: he, him. Women: she, her). Let \mathcal{N}_w denote the number of times a token w appears in a corpus. We define *Male Pronoun Ratio (MPR)* as follows:

$$MPR = \frac{\mathcal{N}_{he} + \mathcal{N}_{him}}{\mathcal{N}_{he} + \mathcal{N}_{him} + \mathcal{N}_{she} + \mathcal{N}_{her}} * 100.$$
 Figure 2 plots *MPR* of our decade-wise movie data sets and contrasts with *MPR* computed using Google Ngrams. Our results indicate that

Probe	BERT _{base}	BERT _{D^{old}_{bolly}}	BERT _{D^{new}_{bolly}}	BERT _{D^{old}_{holly}}	BERT _{D^{new}_{holly}}
cloze ₁	soft, beautiful, pale, tanned, smooth	<u>fair</u> , no, pale, tanned, tan	<u>fair</u> , tanned, golden, smooth, pale	<u>fair</u> , pale, blue, golden, gold	<u>fair</u> , pale, tanned, golden, dark
cloze ₂	woman, man, widow, <u>doctor</u> , <u>slave</u> , <u>soldier</u> , bachelor, <u>merchant</u> , <u>farmer</u> , <u>lawyer</u>	<u>prostitute</u> , <u>servant</u> , woman, <u>slave</u> , bachelor, <u>doctor</u> , <u>lawyer</u> , man, widow, <u>maid</u>	<u>doctor</u> , woman, <u>servant</u> , <u>lawyer</u> , <u>maid</u> , man, <u>nurse</u> , <u>teacher</u> , <u>gardener</u> , lady	woman, <u>slave</u> , <u>servant</u> , <u>nurse</u> , lady, man, <u>teacher</u> , <u>lawyer</u> , peasant, <u>maid</u>	woman, <u>lawyer</u> , <u>doctor</u> , <u>nurse</u> , <u>teacher</u> , man, <u>writer</u> , <u>secretary</u> , <u>prostitute</u> , professional
cloze ₃	man, <u>soldier</u> , gentleman, <u>farmer</u> , <u>merchant</u> , woman, <u>slave</u> , bachelor, <u>doctor</u> , <u>carpenter</u>	man, <u>gentleman</u> , bachelor, <u>lawyer</u> , <u>servant</u> , <u>doctor</u> , <u>farmer</u> , <u>worker</u> , <u>craftsman</u> , <u>slave</u>	<u>doctor</u> , <u>lawyer</u> , <u>policeman</u> , man, <u>farmer</u> , bachelor, <u>gardener</u> , <u>servant</u> , <u>soldier</u> , <u>mechanic</u>	<u>carpenter</u> , <u>police-man</u> , <u>lawyer</u> , <u>soldier</u> , <u>farmer</u> , <u>gentleman</u> , <u>servant</u> , man, peasant, <u>slave</u>	man, <u>lawyer</u> , <u>soldier</u> , <u>doctor</u> , <u>carpenter</u> , gentleman, <u>farmer</u> , <u>clergyman</u> , <u>writer</u> , <u>craftsman</u>

Table 1: Cloze test results. Predicted tokens are ranked by decreasing probability. Positive and negative words are italicized and underlined, respectively. Our completion results using BERT indicate presence of social and gender biases.

even now, both Bollywood and Hollywood exhibit comparable skew in gendered pronoun usage.

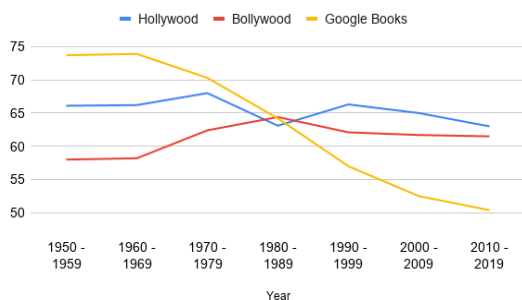


Figure 2: MPR in D_{bolly} and D_{holly}

Cloze Test Using Language Models: When presented with a sentence (or a sentence stem) with a missing word, a cloze task (Taylor 1953) is essentially a fill-in-the-blank task. For instance, in the following cloze task: *In the [MASK], it is very sunny, summer* is a likely completion for the missing word. Given a cloze test, BERT, a well-known language model (Devlin et al. 2019), outputs a series of token ranked by probability. In fact, in the above cloze test, the top three tokens (ranked by probability) predicted by BERT_{base} are: summer, winter and spring. Recent lines of research has explored BERT’s masked query prediction for (1) knowledge base extraction (Petroni et al. 2019) and (2) mining political insights (Palakodety, KhudaBukhsh, and Carbonell 2020).

Following (Palakodety, KhudaBukhsh, and Carbonell 2020), we fine-tune BERT on four sub-corpora: D_{bolly}^{old} , D_{holly}^{old} , D_{bolly}^{new} , and D_{holly}^{new} . We denote the pretrained BERT model as BERT_{base} and a fine-tuned BERT on corpus D as BERT_D. Out of a thorough analysis with several cloze tests with phrase variations, due to paucity of space, in Table 1 we report only three such results for the following cloze tests: (1) *A beautiful woman should have [MASK] skin. (cloze1)*; (2) *A woman should be a [MASK] by occupation. (cloze2)*; and (3) *A man should be a [MASK] by occupation (cloze3)*.

Summary of findings: Given cultural and social bias toward fair skin prevalent in India for years (Karan 2008), it is not

surprising that Bollywood exhibits an affinity toward fairness. However, our analysis indicates Hollywood also shows bias toward lighter skin color. Further, we observe strong presence of gender bias with masked sentences involving men predominantly receiving positive completions in comparison with women (see, Table 1). Future directions include (1) tracking moral sentiment; (2) exploring debiasing techniques; and (3) analyzing representation of other minorities.

References

- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*, 4171–4186.
- Garg, N.; Schiebinger, L.; Jurafsky, D.; and Zou, J. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *PNAS* 115(16): E3635–E3644.
- Karan, K. 2008. Obsessions with fair skin: Color discourses in Indian advertising. *Advertising & society review* 9(2).
- Khan, S.; and Taylor, L. 2018. Gender Policing in Mainstream Hindi Cinema: A Decade of Central Female Characters in Top-Grossing Bollywood Movies. *International Journal of Communication* 12: 22.
- Lewis, M.; and Lupyan, G. 2020. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature human behaviour* 1–8.
- Lison, P.; and Tiedemann, J. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles .
- Madaan, N.; Mehta, S.; Agrawaal, T.; Malhotra, V.; Aggarwal, A.; Gupta, Y.; and Saxena, M. 2018. Analyze, detect and remove gender stereotyping from bollywood movies. In *ACM FAccT*, 92–105.
- Palakodety, S.; KhudaBukhsh, A. R.; and Carbonell, J. G. 2020. Mining Insights from Large-Scale Corpora Using Fine-Tuned Language Models. In *ECAI 2020*, 1890–1897.
- Petroni, F.; Rocktäschel, T.; Riedel, S.; Lewis, P.; Bakhtin, A.; Wu, Y.; and Miller, A. 2019. Language Models as Knowledge Bases? In *EMNLP-IJCNLP*, 2463–2473.
- Taylor, W. L. 1953. “Cloze procedure”: A new tool for measuring readability. *Journalism quarterly* 30(4): 415–433.
- Twenge, J. M.; Campbell, W. K.; and Gentile, B. 2012. Male and female pronoun use in US books reflects women’s status, 1900–2008. *Sex roles* 67(9-10): 488–493.