Evaluation of Gender Bias in Social Media Using Artificial Intelligence

Parthasarathy, Nitya

Gender Bias and stereotyping in human interaction is not uncommon. Sometimes, it could be intentional and sometimes sub-conscious. For example, consider the following description of a man and woman performing similar tasks. “She is a bossy and assertive individual”. Contrast that with “He is a dynamic and confident person”. Indeed, this is could be representative of individual traits. However, when similar gender trait descriptors often repeat and form trends for either gender performing similar roles/tasks, one can then very well associate the specific gender with certain stereotypes. Being aware of both conscious and sub-conscious bias is extremely important as it suggests “conformity into societal roles” as expected behavior. Not only does this encourage bias but it also handicaps individual growth and impedes progress. In our work, we develop Artificial Intelligence(AI) algorithms as well as employ advanced statistical analysis to evaluate gender bias. Though our work readily extend to any class of social media, we limit ourselves to the study of movie reviews available online. First, we studied the distance of gender words to gender typecasts. We demonstrate how close female gender types are to female gender typecasts as compared to male gender typecasts. Next, we adapted a technique using machine intelligence using both Bayesian probabilistic formulas as well as Neural Networks to see if the gender of the sentence could be determined just by looking at surrounding words/role descriptors. Finally, we provide results which shows “fitting of gender with stereotypical adjectives". Our method and algorithms readily extend to also study age/race/social strata related bias and hence is an important advancement in algorithmic bias study in any general social media.

Awards Won:

Fourth Award of $500

American Statistical Association: Certificate of Honorable Mention

University of Arizona: Tuition Scholarship Award