

# Word Embeddings and Gender Stereotypes: Occupation-wise Analysis of the U.S States

ANKITA CHOUDHURY

## ABSTRACT

This study explores the quantitative association between gender stereotypes in occupations expressed through words with empirically observed occupational gender segregation. This analysis is based on Natural Language Processing technique called Word Embeddings. We illustrate how the dynamics in word embeddings help to characterize and quantify stereotypes towards women across select states in the United States of America. Such embeddings trained in a huge corpus of text data of Google News are correlated with Census data to figure out whether they are aligned with demographic and occupational features of some states. It is seen that such embeddings try to capture variation in occupation gender stereotypes for different states in the USA for broad categories of occupations that are either more commonly associated with women (men) like therapist (engineer). At the same time, the study finds that such embeddings have limitations in capturing the male-female segregation at a more disaggregated level either for occupations (say, within engineering) or for states (say, with varying levels of human development). An important finding is that words like manager/engineer have a tendency to show male bias in texts although there is significant improvement in female presence in well-developed states. Such analyses help us understand the bias that are not being captured by words that we use and how policies implemented by states/government can have impact in de-biasing them.

This framework has far-reaching implications as it gives richer insights into cultural associations and categories than possible with prior methods, leading to a successful intersection between Machine Learning and quantitative Social Sciences.

*Keywords:* Word Embeddings, Gender stereotypes, Natural Language Processing

Algorithms, Occupational Segregation