

Critique: “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes”

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou

“Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes” by Nikhil Garg et al. shows how word embedding could be a powerful tool to quantify certain social phenomenon. The framework to compare bias on a gender and ethnic level and to observe historical trend of bias serves as an innovative approach to capture other social problems, such as political division. Abstracting from this paper’s framework, gender and ethnic are on entity level, occupations are a set of characteristics pertaining to the entity. The adjectives used in this study are stereotype related but could extend to any adjective describing the entity.

Surrounding political division, most relevant design is on the stereotype of democrats and republicans. Like gender, we could look at bias in political standing, on the topic of occupation, democrats’ participation in occupation is different from republicans. Literature may capture this real participation rate along with perception of stereotype. Embedding could validate if there is a growing division as perceived. I would expect to see confirmative result of this growing gap. Finding the top associated adjectives could help to better understand the changing image of democrats and republicans. Besides politics, embedding is also helpful for companies to understand their brand image based on the stereotype, as well as identification of competitors’ strategies.

Given embedding could capture attitude changes in response of external events, we apply this use to evaluate the effectiveness of certain movement/policies in history. One external event given in the article is the women’s movement in the 1960s, embedding captures the drastic shift of bias after the movement. To extend the use, take China’s policy on child bearing. From one-child policy to recent encouragement of three-child policy, with embedding, we could evaluate people’s attitude change on babies. Use baby as the entity, providing a list of adjectives and find the most associated during different time periods. From there we could observe if there is a shift attitude corresponding to the policies.

Not limited to above proposed implementation, this paper indeed “opens up a fruitful intersection between machine learning and quantitative social science”.

One minor limitation of the paper is the selection of corpus could be too general. In the example of women’s movement in the 1960s, it may be the many propagandas prevailing in this period that reduced bias in embedding. Though the corpus of Google Books claims to be genre-balanced, use of this corpus may disregard the difference between descriptive observation texts and advocating texts. In such events of strong advocacy, fervent writing may prevail over the realistic depiction. As we could observe from figure 2 (in the research paper), the sharp increase in 1960s faced a downturn in 1970 after the movement; the sharp increase may not be an accurate capture of the prevailing bias in society. Not after the movement, the depiction became more realistic. Therefore, my last proposal is to choose more objective corpus instead of using an all-genre literature of the period.