

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Racial Bias in Emotion Inference: An Experimental Study Using a Word Embedding Method

Permalink

<https://escholarship.org/uc/item/1vj7j16c>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

Authors

Park, Jae Eun

Lee, Yoon Kyung

Hahn, Sowon

Publication Date

2021

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Racial Bias in Emotion Inference: An Experimental Study Using a Word Embedding Method

Jae Eun Park

Seoul National University, Seoul, Korea, Republic of

Yoon Kyung Lee

Seoul National University, Seoul, Korea, Republic of

Sowon Hahn

Seoul National University, Seoul, Korea, Republic of

Abstract

We investigated racial bias in emotion inference by having participants describe the emotion of the featured person in the image, including Asian, Black, and White. We collected 4,197 sentences (63,900 tokens) and used the data to train Word2Vec, a neural network-based word embedding model. We calculated the cosine distance between emotion words and words indicating the target for each racial group in order to measure the strength of association. Although all images portrayed neutral emotions, the results show that negative emotion words were close to Asian and Black, whereas neutral words were close to White. This result indicates that stereotypes contribute to the racial bias in Artificial Intelligence as crowdsourcing workers can generate annotations depending on the featured person's race. Based on the present study, we suggest employing various de-biasing methods such as data augmentation or removing bias components in word embeddings before using Artificial Intelligence in the real world.