THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

DEPARTMENT OF APPLIED MATHEMATICS
應 用 數 學 系

Seminar

Prof Linglong KONG

University of Alberta, Canada

Topic

Debiasing with Sufficient Projection: A General Theoretical Framework for Vector Representations

Date| Time

27th March 2024 (Wednesday) | 10:30 am – 11:30 am (HK Time)

Mode of Delivery

Online via Zoom

Meeting ID| Password

881 0706 7377 | 0327

Zoom Link

https://polyu.zoom.us/j/88107067377?pwd=KUm8dUmbTcYX5LDCzdUqLveOByFjtb.1

Abstract

Pre-trained vector representations in natural language processing often inadvertently encode undesirable social biases. Identifying and removing unwanted biased information from vector representation is an evolving and significant challenge. Our study uniquely addresses this issue from the perspective of statistical independence, proposing a framework for reducing bias by transforming vector representations to an unbiased subspace using sufficient projection. The key to our framework lies in its generality: it adeptly mitigates bias across both debiasing and fairness tasks, and across various vector representation types, including word embeddings and output representations of transformer models. Importantly, we establish the connection between debiasing and fairness, offering theoretical guarantees and elucidating our algorithm's efficacy. Through extensive evaluation of intrinsic and extrinsic metrics, our method achieves superior performance in bias reduction while maintaining high task performance, and offers superior computational efficiency.

ALL ARE WELCOME