THE HONG KONG POLYTECHNIC UNIVERSITY
DEPARTMENT OF MANAGEMENT AND MARKETING

Departmental Research Seminar

Relieving Racial Bias in Hate Speech Detection with a Small Number of Expert Annotations

By

Prof. Michael C.L. Chau
University of Hong Kong

Date : 14 Oct 2024 (Mon)
Time : 10:30 am – 12:00 noon
Venue : M802, PolyU

Abstract

Hate speech is a major problem on social media platforms. Automatic hate speech detection methods relying on machine learning models, which learn from manually labeled datasets, have been proposed in both academia and industry. However, there is increasing evidence that hate speech detection datasets labeled by general annotators (e.g., amateurs or MTurk workers) contain systematic bias, as they cannot effectively consider language use differences among different speakers. When such biased datasets are used to train machine learning models, the resulting models will also be biased. Unlike general annotators, experts can produce much less biased annotations. However, expert annotations cannot be efficiently obtained in large quantity. This paper bridges the gap by adopting a weakly supervised learning method for hate speech detection using a small number of expert annotations. We propose a novel design that uses contrastive learning and prompt-based learning based on large language models, incorporating a group estimator, a pair generator, and external knowledge. Using real-world Twitter posts written by African American English speakers and other racial groups as an example, extensive experiments were conducted to demonstrate the superior performance of the proposed method. The proposed approach was also evaluated on data in the LGBTQ+ community and achieved consistent results. The study has important academic and practical implications for hate speech detection and large language models.

Prof. Michael C.L. Chau is a Professor in Innovation and Information Management in the HKU Business School at the University of Hong Kong. He served as the Warden of Lee Chi Hung Hall (2009-2021) and the Program Director/Coordinator of the BBA (Information Systems) program (2006-2009, 2012-2018). He is also an Honorary Fellow of the HKU-HKJC Centre for Suicide Research and Prevention. He received a Ph.D. degree in Management Information Systems from the University of Arizona and a B.Sc. degree in Computer Science (Information Systems) from the University of Hong Kong. His research interests include business analytics, artificial intelligence, web mining and social media, electronic commerce, fintech, smart health, security informatics, human-computer interaction, and IT in education.

He has published more than 150 articles in premier journals and conferences in information systems, computer science, and information science. He has received multiple international research awards and has been highly ranked in several research productivity studies.

Prof. Chau has been active in serving the research community. He is a member of the AIS College of Senior Scholars and the Program Co-chair of PACIS 2024 and ICIS 2013. He has served on the organization committee and program committee of many information systems and computer science conferences, as well as the editorial board of multiple journals. He is a founding co-chair of the Pacific-Asia Workshop on Intelligence and Security Informatics (PAISI 2006-2019).

All interested are welcome.

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Department of MANAGEMENT & MARKETING
管理及市場學系