Utterance-Level Incongruity Learning Network for Multimodal Sarcasm Detection

Liujing Song*†, Zefang Zhao*†, Yuxiang Ma§, Yuyang Liu¶ and Jun Li*‡
*Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
†University of Chinese Academy of Sciences, Beijing, China
§School of Computer and Information Engineering, Henan University, Kaifeng, China
¶Institute of Medical Information, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
{songliujing, zhaozefang, lijun}@cnic.cn, y.x.ma@hotmail.com, liu.yuyang@imicams.ac.cn, ‡Corresponding Author

***Abstract*—With the exponential growth of user-generated online videos, multimodal sarcasm detection has recently attracted widespread attention. Despite making significant progress, there are still two main challenges: 1) previous works primarily relied on word-level feature interactions to establish relationships between inter-modality and intra-modality, which could potentially lead to the loss of fundamental emotional information. 2) they obtained the incongruity information only interacted with textual modality, which may lead to the neglect of incongruities. To address these challenges, we propose a novel utterance-level incongruity learning network (ULIL) for multimodal sarcasm detection, where the multimodal utterance-level attention (M-ULA) and incongruity learning network (ILN) are the two core modules. First, we present M-ULA to interact with utterance-level multimodal information, complementing word-level features. Furthermore, ILN selects primary modality and auxiliary modality automatically, and leverages cross-attention and self-attention to learning incongruity representations. We conduct extensive experiments on public datasets, and the results indicate that our proposed model achieves state-of-the-art performance in multimodal sarcasm detection.**

***Keyword*—multimodal sarcasm detection, utterance-level attention, incongruity learning**

liujing Song received his Bachelor's degree from Zhengzhou University. Now she is a Ph.D. student in the University of Chinese Academy of Sciences. His research focuses on natural language processing and data mining.

Zefang Zhao received his Bachelor's degree from Taiyuan University of Technology. Now he is a Ph.D. student in the University of Chinese Academy of Sciences. His research focuses on deep learning, natural language process and sentiment analysis.

Yuxiang Ma is currently an associate professor in the School of Computer and Information Engineering at Henan University. He received the B.S. degree from Henan University in 2013, and the Ph.D. degree from the Computer Network Information Center, Chinese Academy of Sciences in 2019. His main research interests include network security, mobile computing, and privacy enhancement technologies.

Yuyang Liu received the Ph.D. degree from University of Chinese Academy of Sciences. He is currently a Research Associate Professor in Institute of Medical Information, Chinese Academy of Medical Sciences. His research interests include complex medicial knowledge networks, clinical decision support system, and various data mining and artificial intelligence applications across medical informatics.

Jun Li is a research fellow and doctoral supervisor at the Computer Network Information Center of Chinese Academy of Sciences, specially appointed researcher of Chinese Academy of Sciences. His main research interests are artificial intelligence and big data technical applications and future Internet architecture.