

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Ms. YANG Ruichao
Date	16 October 2024 (Wednesday)
Time:	1:30 pm – 3:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	 DLB 637, 6/F, David C Lam Building, Shaw Campus ZOOM (Meeting ID: 988 7719 6375) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 12:00 nn, 15 October 2024)

Detecting Online Misinformation and Stance with Weakly Supervised Models Assisted by Contextualized Social Wisdom

<u>Abstract</u>

In recent years, people have witnessed a proliferation of misinformation, such as rumors and fake news, spreading rampantly on social media and startling the public. Recent studies have shown that rumor verification and stance detection are two relevant tasks that can complement each other despite their inherent differences. For instance, coarse-level claim rumors can be debunked by cross-checking the related posts' stances, and fine-grained level stances are influenced by the nature of the rumor. However, many stance detection methods require annotations at the post level for training, which can be costly and labor-intensive to obtain. Fake news debunking primarily focuses on determining the truthfulness of news articles, which oversimplifies the issue as fake news often combines elements of both truth and falsehood. Thus, it becomes crucial to identify fine-grained level instances of misinformation within the coarse-level articles. Besides, one of the challenges also lies in the absence of a training dataset with sentence-level annotations regarding veracity. To address the issue of annotation scarcity, this thesis focuses on weakly supervised learning for online misinformation and stance detection tasks: joint rumor verification and stance detection, as well as fake news detection.

Enlightened by multiple instance learning (MIL), we transform the two multi-class problems into several binary classification tasks, and then we propose a novel MIL-based method for joint rumor verification and stance detection. After that, we introduce a novel reinforcement tuning framework for joint rumor verification and stance detection via weakly supervised learning with large language models (LLMs). Besides, we investigate a novel task in the field of fake news debunking, which involves detecting sentence-level (fine-grained level) misinformation. We also conduct extensive experiments on various datasets coming from different platforms, and the experimental results demonstrate the promising performance of the methods proposed in this thesis for online misinformation and stance detection tasks.

*** ALL INTERESTED ARE WELCOME ***