Research Statement

Lo Siaw Ling

School of Computing and Information Systems, Singapore Management University
tel: (65) 6828-0287; *email:* sllo@smu.edu.sg
21-Dec-2023

Background

Is pretrained models and deep learning the answer to extracting useful insights from informal textual data (e.g., social media)? Even though there are many promising results from pretrained models and deep learning, we have observed that a pretrained model that is trained on one event cannot be generalized to extract situational awareness tweets from an unseen event. It is essential to perform further testing on a promising classifier that is built using a pretrained model (be it from static or contextualized word embeddings) to ensure that the model can generalize well for various events before it is deployed for any real-world application. On top of that, it is not always necessary to deploy the latest or the state-of-the-art methods, if there is no obvious improvement in the performance. To have a real-world impact, the focus should be on the quality and relevancy of the data or engage a data-centric approach, instead of merely improving on various algorithms.

Research Areas

Actionable situational intelligence for urban events

This project was funded by ST Engineering Mission Software & Services Pte Ltd (*$865,569 for 2 years – May 2021 to Apr 2023*), to provide sense-making capability on an urban event based on social media content, which can be an incident of an armed assault or a crisis of a sudden riot. The research focuses on in-depth analysis of the event including timeline-based situational and emotional changes and relationship among the key-entities. The following are some of the research topics:

1) Data-centric approach in preparing data to build a more generalized AI model
2) Hybrid approach combining knowledge-based/knowledge graph and machine-learning/deep learning (including domain adaptation and transfer learning) in training AI models
3) Incorporate common sense knowledge in understanding the intent of persons and emotion involved in the event in order to uncover key relationship and information summarization

Outcome:
1. Project Milestone Delivery (*all four milestones delivered on time*)
    a. The first three milestones (D1, D2, D3) delivered multiple AI models that identify various event trigger types by leveraging sequence tagging approach (BERT-BiLSTM-CRF model) to detect multiple event types on a same tweet. A Localized knowledge base (KB) is constructed to organize location information of Singapore to further improve the essential information extraction. In order to have a closer-to-real-world evaluation, we have collected and annotated a separate test dataset consisting of new events unseen during the training to assess the model developed. The team has developed a data-centric approach using "difficult negative" training data that helps to achieve the best performance in unseen events.
    b. Fourth and Last milestone (D4, April 2023): In order to capture more detailed and relevant situational awareness information, we leverage knowledge base and graph,

including the localized knowledge base, related Singlish dictionary and common-sense evaluation. The scope of D4 is to integrate with the D3 released AI model to extract timeline-based 1) situational summary, details and entities, and 2) emotional details. The model detects events from tweets collected in batch mode. For the first task, we have included more representative events in the training data. Besides that, the AI model has been improved using multi-task learning to handle both crisis details and crisis impact sequence labelling, with BERTweet enhancements. The second task implements a multi-task learning model with attention mechanism that predicts sentiment and emotion together with representative phrases extraction. We have also incorporated intent analysis using common sense knowledge base and COMET-ATOMIC 2020 BART model in this delivery.

2. The overall project: Artificial Intelligence Emergency Sensemaking Platform (which includes the modules developed in SMU) was nominated for the Institution of Engineers, Singapore (IES) *Prestigious Engineering Achievement Awards 2023* under the Applied Research and Development Category

3. One HICSS 2023 full paper (*nominated for best paper award*), one PACIS 2023 full paper, one ICIS 2023 full paper, one revised paper to *Information Processing and Management* journal, and one working paper on the research findings (*details in page 3*).

AI in Education

With the release of ChatGPT, Generative AI has impacted all industries, including education. Although it sounds very promising and has been shown to improve productivity of users, there are known problems, such as hallucination. It is of interest to leverage Generative AI or more specifically, Large Language Models (LLMs), for teaching and learning. This combined with text analytics and previously studied doubt identification research have opened more research opportunities. For example, classifying users' comments on online programming support platform, StackOverflow, based on presence of doubts can provides insights to expert users to address the concern and elevate the quality of the answers.

Outcome:
1. Awarded *SMU Educational Research Fellowship 2023* on project "DE.AI: Debunking misconceptions for Enhanced Student Learning with the Power of Generative AI"
2. One full paper in PACIS 2023 (*details in page 3*).
3. co-PI for submitted MOE-TRF – "PromptTutor - Generative AI-enabled Personalised Tutor for Promoting Reflection and Learning in Programming Courses"
4. The work has been expanded into a collaboration with Pei Hua Cher, Assistant Professor (AI & Learning Analytics) and Prof Fernanda Bello, Professor of Technology Enhanced Learning and Innovation of Duke-NUS to research on the use of AI in reflection study.

Publications and Outputs:

Conference papers

1. <u>LO Siaw Ling</u>, LEE Kahhe, ZHANG Yuhao. 'Is a Pretrained Model the Answer to Situational Awareness Detection on Social Media?', in *Proceedings of the 56th Hawaii International Conference on System Sciences*, 2023, pp. 2110-2119 (*nominated for best paper award)*

2. Zhang Yuhao, <u>Lo Siaw Ling</u>, Win Myint, Phyo Yi. 'Impact of Difficult Negative on Twitter Crisis Detection', in *Proceedings of 2023 Annual Pacific Asia Conference on Information Systems*, 2023, 156. https://aisel.aisnet.org/pacis2023/156

3. CHEN Tianhao, OUH Eng Lieh, TAN Kar Way, <u>LO Siaw Ling</u>. 'Machine-learning approach to automated doubt identification on stack overflow comments to guide programming learners', in *Proceedings of 2023 Annual Pacific Asia Conference on Information Systems*, 2023, 135. https://aisel.aisnet.org/pacis2023/135

4. Zhang Yuhao, <u>Lo Siaw Ling</u>, Win Myint, Phyo Yi. 'Transformer-Based Multi-Task Learning for Crisis Actionability Extraction', in *Proceedings of 2023 Annual International Conference on Information Systems*, 2023, 1. https://aisel.aisnet.org/icis2023/socmedia_digcollab/socmedia_digcollab/1

Teaching Cases

5. <u>LO Siaw Ling</u>, NG Graham, LIM Thomas, 'Blazing a Trail in the Charity Sector: Singapore's Metta Welfare Association Raises Funds by Selling NFTs*'* in *Harvard Business Publishing* on 16 August 2023. https://hbsp.harvard.edu/product/052SMU-PDF-ENG

Paper submitted and revised:

1. 'Unveiling the Dynamics of Crisis Events: Sentiment and Emotion Analysis via Multi-Task Learning with Attention Mechanism and Subject-based Intent Prediction' to *Information Processing and Management* journal - Phyo Yi Win Myint, <u>Lo Siaw Ling</u>, Zhang Yuhao (revised submitted on 25 Oct 2023)

Paper in progress:

1. 'Real-World Actionable Information Extraction System (AIES) for Social Media' – Zhang Yuhao, Lo Siaw Ling, Phyo Yi Win Myint – target to submit to Decision Support Systems (Journal)
2. 'Reflection in education using AI and learning analytics – an integrative review' – Cher Pei Hua, Lo Siaw Ling, Fernando Bello Guerra – target to submit to Computers and Education: Artificial Intelligence (Journal)
3. 'Leveraging LLMs and misconceptions for automatic MCQ generation' -target to submit to AIED conference – Lo Siaw Ling, Cheryl Goh Sze Pei

New research opportunity:

Submitted a 2-year research proposal to DSO (titled *Human Performance Assessment using Semantic-based Prediction Methods*) in developing an AI model to assess the performance of pilot trainees. This project will explore three tasks:

1. **Semantic segmentation of multimodal user activity data**: In this task, we would like to learn semantic labels to multiple streams of user activity data as a user performs an assessment task. Each semantic label should not only provide meaning to the data, but also allow the assessment results to be predicted with higher accuracy.
2. **Performance assessment prediction with semantic segment features**: To address this task, we propose a Semantic Time-Series Encoder to summarize a user's activity and segment data into a representation vector which can be directly used for performance prediction using a Semantic Time-Series Classifier or Semantic Time-Series Regressor.
3. **Explainable assessment prediction**: To aid the understanding of performance prediction outcome, we also propose to investigate the possibility of an Assessment Explanation Model to generate a semantic explanation of the outcome of a user assessment based on the results of our proposed Semantic Time-Series Classifier or Regressor.