Research Statement

LIM Ee Peng
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-0781; Email: eplim@smu.edu.sg
19 (Day) 11 (Month) 2023 (Year)

Background

My research interests cover data mining, machine learning, and information retrieval. In my research, I would like to develop novel methods that model and analyse user generated structured and unstructured data for the purpose of extracting knowledge embedded in the data and improving social and urban experience through recommendation and prediction. In my research, I venture beyond standard datasets and study new application domains. I believe these new domains will be able to offer interesting datasets and multidisciplinary research problems.

Other than pursuing my research interests, I have spent significant time and efforts on grant proposal writing which includes my own grant proposals (1 MOE Tier 2 grant, 1 NMRC grant, and 1 DSO grant) as well as proposals in collaboration with other faculty members (1 MOE Tier 3, 1 MOE Tier 2, and 1 Science of Learning). I am optimistic that some of them will bring in new resources to support my works.

In the following, I summarize my recent works in a few research areas.

Research Areas

In **knowledge graph-based information retrieval**, my PhD student and I have studied *Contextual Path Generation* (CPG) which refers to the task of generating knowledge path(s) between a pair of entities mentioned in an input textual context to determine the semantic connection between them. Such knowledge paths, also called contextual paths, can be very useful in many advanced information retrieval applications. To perform CPG task well, one has to address its three challenges, namely path relevance, incomplete knowledge graph, and path well-formedness. This paper designs a two-stage framework the comprising of the following: (1) a knowledge-enabled embedding matching and learning-to-rank with multi-head self-attention context extractor to determine a set of context entities relevant to both the query entities and context document, and (2) a non-monotonic path generation method with pretrained transformer to generate high quality contextual paths. Our experiment results on two real-world datasets show that our best performing CPG model successfully recovers 84.13% of ground truth contextual paths, outperforming the context window baselines. Finally, we demonstrate that non-monotonic model generates more well-formed paths compared to the monotonic counterpart.

In **financial forecasting**, I have guided my PhD student, Gary Ang, to study investment and risk management using network modeling. Stock price movements in financial markets are influenced by large volumes of news from diverse sources on the web, e.g., online news outlets, blogs, social media. Extracting useful information from online news for financial tasks, e.g., forecasting stock returns or risks, is, however, challenging due to the low signal-to-noise ratios of such online information. Assessing the relevance of each news article to the price movements of individual stocks is also difficult, even for human experts. In this article, we propose the Guided Global-Local Attention-based Multimodal Heterogeneous Network (GLAM) model, which comprises novel attention-based mechanisms for multimodal sequential and graph encoding, a guided learning strategy, and a multitask training objective. GLAM uses multimodal information, heterogeneous relationships between companies and

leverages significant local responses of individual stock prices to online news to extract useful information from diverse global online news relevant to individual stocks for multiple forecasting tasks. Our extensive experiments with multiple datasets show that GLAM outperforms other state-of-the-art models on multiple forecasting tasks and investment and risk management application case-studies.

In **recommender systems research**, I focus on developing next basket recommendation. Next-basket recommendation (NBR) is a recommendation task that predicts a basket or a set of items a user is likely to adopt next based on his/her history of basket adoption sequences. It enables a wide range of novel applications and services from predicting next basket of items for grocery shopping to recommending food items a user is likely to consume together in the next meal. Even though much progress has been made in the algorithmic NBR research over the years, little research has been done to broaden knowledge about the evaluation of NBR methods, which is largely based on the offline evaluation experiments and binary relevance paradigm. Specifically, we argue that recommended baskets which are more similar to ground truth baskets are better recommendations than those that share little resemblance to the ground truth, and therefore, they should be granted some partial credits. Based on this notion of non-binary relevance assessment, we propose new evaluation metrics for NBR by adapting and extending similarity metrics from natural language processing (NLP) and text classification research. To validate the proposed metrics, we conducted two user studies on the next-meal food recommendation using numerous state-of-the-art NBR methods in both online and offline evaluation settings. Our findings show that the offline performance assessment based on the proposed non-binary evaluation metrics is more representative of the online evaluation performance than that of the standard evaluation metrics.

Future Research. I shall continue using my research expertise in data mining, information retrieval and AI to explore more varied research problems in novel application domains. I will also expand my collaboration with domain experts to identify and work on important social and urban research problems.

Selected Publications and Outputs

Pei-Chi Lo and Ee-Peng Lim. 2023. Non-Monotonic Generation of Knowledge Paths for Context Understanding. ACM Trans. Manage. Inf. Syst. Just Accepted (October 2023)

Gary Ang and Ee-Peng Lim. 2023. Investment and Risk Management with Online News and Heterogeneous Networks. ACM Trans. Web 17, 2, Article 8 (May 2023), 24 pages.

Liu, Y., Achananuparp, P. & Lim, EP. Non-binary evaluation of next-basket food recommendation. User Model User-Adapted Interaction (2023).