香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr Lei LI
Date	27 April 2022 (Wednesday)
Time:	10:30 am – 12:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	Zoom ID: 941 3035 0129 (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/sem-zm (Deadline: 6:00 pm, 26 April 2022)

Natural Language Explanation for Recommendations and Beyond

Abstract

Explainable artificial intelligence (XAI) could unveil the decisions made by AI, help human understand machine behavior, and further build trust between end users and AI systems. However, most existing XAI approaches are primarily mathematical, and also specialized to domain experts. Although they can help researchers understand the underlying working mechanism of AI models, the explanations are hardly comprehensible to ordinary users. As a user-centric application, recommender systems interact with and serve a great number of such users. As a consequence, how to explain recommended products to them, in order to help them make informed choices and provide better service, becomes a critical and practical problem.

As a primary media of communication, language can be easily understood by almost everyone. Therefore, natural language explanation for recommendations has gained increasing attention recently. Despite of that, little work has been done to provide explanations from the perspective of a user's changing context, such as companion and destination if the recommendation is a hotel. To fill this research gap, we have devised a new context-aware recommendation approach that particularly matches latent features to explicit contextual features for producing context-aware explanations. But the explanation format in this approach, as with many previous works, is limited to predefined templates, which could restrict explanation expressiveness, and thus may not be able to well explain the specialty of a recommendation. In an attempt to further enrich explanation expressiveness and quality, we have proposed a neural template generation approach that can learn templates from data. In order to generate such template-like explanations, an item feature must be specified in advance, either automatically or manually. To accommodate situations where features are unavailable, we have designed a more general method that can generate natural language explanations with or without features.

The proceeding two methods can generate high-quality explanations, but do not always guarantee factual correctness, e.g., "four-horned unicorns" as produced by a well-known pre-trained language model. To cope with this problem, we wonder whether explanations for a recommendation could be ranked, as if they are web pages returned by a search engine in accordance with a given query, saving the need to worry about the content of explanations. To enable such explanation ranking, we have created benchmark datasets by automatically identifying nearly identical sentences across different user reviews, based on the wisdom of the crowd. In addition, the ranking formulation makes it possible for standard evaluation of explanations via ranking-oriented metrics. Based on this, we have studied if purposely selecting some explanations could reach certain goals, e.g., improving recommendation accuracy. This could potentially lead to unfaithful explanations that attempt to lure users' clicking and purchasing. Hence, at the end of this thesis, we discuss this type of unintentionally negative effects as well as other open issues in explainable recommendations, such as bias and fairness.

We believe that these works conducted in this thesis are non-trivial and meaningful to the community of XAI. They are instantiated on the scenario of explainable recommendations, but could be generalized to a broader scope of fields in AI, e.g., dialogue systems.

***** ALL INTERESTED ARE WELCOME *****