University of Pittsburgh | School of Computing and Information

Dissertation Defense

Doctor of Philosophy in Computer Science

"**Domain Robustness in Multi-modality Learning and Visual Question Answering**" by **Mingda Zhang**

Date: November 30, 2021
Time: 12:00pm – 2:00pm
Place: https://pitt.co1.qualtrics.com/jfe/form/SV_1SuRnfw2MzR5GYu

Committee:

- Dr. Adriana Kovashka, Assistant Professor, Department of Computer Science, School of Computing and Information
- Dr. Rebecca Hwa, Professor, Department of Computer Science, School of Computing and Information
- Dr. Diane Litman, Professor, Department of Computer Science, School of Computing and Information
- Dr. Seong Jae Hwang, Assistant Professor, Department of Computer Science, School of Computing and Information
- Dr. Daqing He, Professor, Department of Informatics and Networked Systems, School of Computing and Information

Abstract:
Humans perceive the world via multiple modalities, as information from a single modality is usually partial and incomplete. This observation motivated the development of machine learning algorithms capable of handling multi-modal data and performing more intelligent reasoning. The recent resurgence of deep learning brings both opportunities and challenges to multi-modal reasoning tasks. On one hand, its strong representation learning capability provided a unified approach to represent information across multiple modalities. On the other hand, properly training such models typically requires enormous data, which is not always feasible especially for the multi-modal setting.

One promising direction to mitigate the lack of data for training deep learning models is to transfer knowledge (e.g., models gained from solving related problems) to low-resource domains. This procedure is known as domain adaptation and has demonstrated great success in various visual and linguistic applications. However, how to effectively transfer knowledge in a multi-modality setting remains an open question. We chose multi-modal reasoning as our target task and aimed at improving the performance of deep neural networks on low-resource domains via domain adaptation. We first briefly discussed our prior work about advertisement understanding (as a typical multi-modal reasoning problem) and shared our experience from addressing the data-availability challenge. Next, we turned to visual question answering, a more general problem that involves more complicated reasoning. We evaluated mainstream models and classic single-modal domain adaptation strategies and showed that existing methods usually suffered significant performance degradation when directly applied to multi-modal setting. Lastly, we studied domain adaptation approaches with different supervisions (e.g., unsupervised, self-supervised, semi-supervised or fully supervised training) and shared the key components for improving domain robustness in visual question answering.