DISSERTATION DEFENSE
DEFENSE

Automating Biomedical Abstract Screening using Network Embedding

Eric Lee
Emory University

Abstract: Systematic review (SR) is an essential process to identify, evaluate, and summarize the findings of all relevant individual studies concerning health-related questions. However, conducting a SR is labor-intensive, as identifying relevant studies is a daunting process that entails multiple researchers screening thousands of articles for relevance. Automating evidence synthesis using machine learning models has been proposed but primarily focuses on the text and ignores additional features like citation information. Recent work demonstrated that citation embeddings can outperform the text itself, suggesting that better network representation may expedite SRs. Yet, how to utilize the rich information in heterogeneous information networks for network embeddings is understudied. Also, the lack of a unified source that includes the metadata of biomedical literature makes the research more challenging. To deal with this problem, we propose four works. First, we propose a model that exploits three representations, documents, topics, and citation networks to show the effectiveness of the additional features. Second, we introduce the PubMed Graph Benchmark, one of the largest heterogeneous networks to date, which aggregates the rich metadata into a unified source that includes abstracts, authors, citations, MeSH terms, etc. Third, we propose a heterogeneous network embedding model that uses a community-based multi-view graph convolutional network for learning better embeddings for evidence synthesis. Lastly, we propose a hierarchical network embedding model that uses the Poincare embedding model on the MeSH tree and the citation network.

Thursday, September 28, 2023, 1:30 pm
Mathematics and Science Center: MSC E406

Advisor: Joyce Ho

COMPUTER SCIENCE
EMORY UNIVERSITY