香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. Kejing YIN
Date	July 6, 2021 (Tuesday)
Time:	9:00 am – 11:00 am (35 mins presentation and 15 mins Q & A)
Venue:	Zoom ID: 979 5937 1071 (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/sem-zm (Deadline: 1:00 pm, July 5, 2021)

Learning Phenotypes from Electronic Health Records using Robust Temporal Tensor Factorization

Abstract

Computational phenotyping — discovering meaningful combinations of clinical items, e.g. diagnosis and medications, from the increasingly accessible electronic health records (EHR) data has shown effective to characterize complex health conditions of patients. However, its robustness against low data quality can be greatly compromised due to open challenges including missing inter-modal interactions, temporal irregularity, heavy missingness, etc. In this thesis research study, we propose a series of robust tensor factorization models to address these issues.

First, we propose a hidden interaction tensor factorization (HITF) model and a collective HITF framework to recover the unrecorded inter-modal correspondence jointly with the learning of latent phenotypes. Our results show that discovering hidden interactions significantly improves the quality of the phenotypes discovered and the downstream prediction tasks.

Second, we propose a collective non-negative tensor factorization (CNTF) model to extract phenotypes from temporally irregular EHR data. In particular, we represent the data of each patient using an individual tensor to preserve the temporal information and handle the temporal irregularity. The results show that the phenotypes that appear at different stages of the disease progression can be effectively separated.

Third, we propose a temporally dependent PARAFAC2 factorization (TedPar) model to further capture the temporal dependency between phenotypes by capturing the transitions between them over time. Our evaluation shows that modeling the temporal transition between phenotypes helps improve generalization to unseen patients.

Forth, we propose a logistic PARAFAC2 factorization (LogPar) model to jointly complete the one-class missing data in the binary irregular tensor and learn phenotypes from it. In particular, we extend the PU learning technique to the setting of irregular tensor factorization. The results show that the resulting model is significantly more robust to the missingness in binary data.

Finally, we extend our analysis to continuous time series by proposing the context-aware time series imputation (CATSI) to capture the overall health condition of patients and use it to guide the imputation of clinical time series. Extensive experiments show that it outperforms existing methods.

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