Computer Science
Defense

Temporal Irregular Tensor Factorization and Prediction for Health Data Analysis

Yifei Ren
Emory University

Abstract: Tensors are a popular algebraic structure for a wide range of applications, due to their exceptional capability to model multidimensional relationships of the data. Among them, regular tensors with aligned dimensions for all modes have been extensively studied, for which various tensor factorization structures are proposed depending on the applications. However, regular tensor decomposition is incapable of handling many real-world cases involving time, due to its irregularity. Electronic health records (EHRs) are often generated and collected across a large number of patients featuring distinctive medical conditions and clinical progress over a long period of time, which results in unaligned records along the time dimension. Recently, PARAFAC2 has been re-popularized for successfully extracting meaningful medical concepts (phenotypes) from such temporal EHR by irregular tensor factorization. ¡br¿ However, existing PARAFAC2 methods suffer from three major limitations that impact their applicability to practical temporal EHR data analysis: 1) they are not robust to missing and erroneous elements in the data; 2) they fail to model the non-linear temporal dependency of patients' disease states, and are designed only for a single data type – numeric or binary; 3) they are completely unsupervised, i.e., they attempt to learn the latent factors to best recover the original observations without considering downstream predictive tasks. While there are models that use extracted phenotypes for predictive tasks, they are trained separately and only consider a single prediction task, which ignores auxiliary information from other predictive tasks. ¡br¿ To address these limitations, we make three main contributions in this dissertation. We first propose a robust PARAFAC2 tensor factorization method for irregular tensors with a new low-rank regularization function to handle potentially missing and erroneous entries in the input data. We then propose a generalized, low-rank Recurrent Neural Network (RNN) regularized robust irregular tensor factorization for more accurate temporal modeling, with the flexibiity to choose from different losses to best suit different types of data in practice. Finally, we propose a supervised irregular tensor factorization framework with multi-task learning for joint optimization of phenotype extraction and predictive learning, which can yield not only more meaningful phenotypes but also better predictive accuracy. ¡br¿ **Join Zoom Meeting** Venue: https://emory.zoom.us/j/2878070037

Thursday, December 2, 2021, 12:00 pm
https://emory.zoom.us/j/2878070037

Computer Science
Emory University