An Interpretable Deep Learning Framework for Disease Staging

Liuyi Yao¹, Zijun Yao, PhD², Jianying Hu, Ph.D², Zhaonan Sun, PhD²
¹University at Buffalo, Buffalo, NY; ²IBM Research, Yorktown Heights, NY

Introduction

Disease staging is a clinically based measure of severity that uses objective medical criteria to assess the state of disease progression¹. Subdivision of disease stages allows more precise classification. Disease staging and sub-staging have valuable applications in many areas such as assessing the quality of care, analyzing clinical outcomes, reviewing of utilization of resources, etc. Data-driven disease stages and sub-stages learning with observational data is challenging. Moreover, it is crucial to generate clinically meaningful descriptions for the learned stages. We propose an interpretable deep learning framework, named Deep Staging, for data-driven disease staging and sub-staging, as well as generating clinically relevant characterizations of the learned disease stages and sub-stages.

Method and Materials

We assume that a target disease can be separated into a number of disease stages, and the number of disease stage is given. The goal is to learn disease stages and their clinical description using longitudinal observations (Xt). The proposed framework consists of two steps (Figure 1), which are the Deep Staging model (left panel) and the post-modeling description learning (middle panel). The Deep Staging model is a Recurrent Neural Network (RNN) model that characterize the temporal patterns. The RNN model consists of two key components: 1) generating disease stage probability distribution (Pt) and 2) reconstructing observations (Xt). In the next step, the description learning leverages the two key components and learns representations of each disease stage in the clinic observation space. In this way, medical practitioners can directly understand the characteristics of the disease stages w.r.t. observable clinical measures or manifestations, without having to digging into the complicated RNN model. The framework also provides the flexibility for users to divide the learned disease stages into sub-stages with pre-defined topological structures.

Figure 1. Left: architecture of the Deep Staging model; Middle: description learning process; Right: Distribution of a group of most frequent ICD-9 codes in four states.

A preliminary experiment was performed on a Chronic Kidney Disease (CKD) cohort from a real-world EHR database with the number of stages set to 4. The diagnosis coded in ICD-9 codes were used as the input. The right panel of Figure 1 shows the probabilities of a group of ICD-9 codes in the learned stages. The results showed that the distributions of some critical ICD-9 codes differ in the four stages, indicating that the framework is capable of separating different typical status along the progression pathway. The probabilities also provide a clinical description of each learned stages in terms of the most common diagnosis in the corresponding stage.

Conclusion

This article proposed a deep learning framework for disease staging. It also generates stage characterization in the original input space. Therefore, users can directly understand the learned stages through observable measures. More thorough experiments will be conducted in future for better investigating the proposed framework.

References