



Week 2 Reflections

| | |
|----------------|---------------------------------------|
| ☰ Course | CS 598 - Deep Learning for Healthcare |
| ☰ Submitted By | Rohit Bhagwat (rohitb2@illinois.edu) |

General questions

What are the most useful things you learned from this chapter or video?

The chapter and the videos provided a very detailed introduction to Health Data. In Week 1, we had to undergo the CITI program module training which gave a brief history of using health data for research and the importance of standards and compliance in this area.

The chapter provided summary on the various types of Health Data - Structured Health Data, Unstructured Clinical Notes, Continuous signals, Medical Imaging Data and Biomedical data for silico drug discovery. (Honestly before I started this course, I thought the Health Data is limited to Patient data.) It was also very helpful to understand the significance of Electronic Health Record (EHR) adoption and how it can help to increase the quality of life or the life span of humans.

I was also very fascinated to learn about the standards used to document the data. For example - Diagnosis details can be represented using ICD codes, Procedures can be represented using CPT codes, Drugs can be represented using NDC codes. Understanding these codes clearly plays an important role in modeling health care data.

What are the typos in this chapter?

Some typos from the PDF -

- Page 14:“Objective part presents the objective findings such as lab test results and imaging rests (tests)”
- Page 16: “it is important to provide more convincing evidence and interpretable explanation(explanations)”

What improvements do you want to see in this chapter?

In my humble opinion the PDF document can possibly contain some more diagrams to engage the readers, especially when discussing Health Data Standards - similar to what we see on the Slides.

I'd also love to read more about the history of how these standards were created (although, this can be an optional reading).

Questions specific to this health data chapter

What are the most useful health data for predicting patient outcome (e.g., mortality)?

In my opinion, EHR data is probably the most useful data for predicting patient outcomes. EHR data is structured and includes important and helpful information like patient's symptoms, the diagnosis and also includes past medications or procedures performed on the patient. This data from the past visits can be used to predict future events. Discovering temporal relations is crucial for making an accurate assessment or prediction of any health. Health outcomes research can use this data to identify the links to the conditions, treatments and the outcomes of the past and predict the future outcomes.

What are the most accessible health data? And why?

The publicly available datasets such as medical literature (PubMed), Medical Ontologies (SNOMED) and drug data (DrugBank) are easily accessible. Additionally some medical imaging data is also accessible, the internet and the big data revolution has made this data accessible for research.

What are the most difficult health data (to access and to model)?

EHR & Clinical Notes. This is due to privacy concerns as it contains highly sensitive data and PHI. As a result, very limited clinical notes are openly accessible for method development. Additionally, it is likely hard to model well because it is unstructured and includes a lot of abbreviations and potential misspellings. There's a large amount of medical knowledge encoded in the text - such as medical literature and clinical guidelines.

The above challenges make it extremely difficult to access and model the data.

What are the types of health data that you want to add to this chapter?

I could think of a couple of datasets, or examples -

- Health outcomes data
- Commercial and 'structured' data from pharmacies