

CS410 Fall 2022
Cristina Ross
NetID: cb37
Group Name: The Ross Factor

Course Project Progress Report

In the project proposal six separate tasks were listed (1) Build collection of notes (2) Build background model (3) Build out unique background model generation (4) Build functionality to identify unique topics (5) Build ML model and (6) Evaluate model. So far in the project the majority of time has been spent attempting to find a publicly available dataset of healthcare notes with a certain percentage that have a unique medical diagnosis. So far few have been found. The first consists of Historical and Physical notes¹, with six total notes, two of which admit with chest pain. This could be a starting point with 30% prevalence of chest pain, but six notes is not a sufficient amount for tasks 2-6. The second source had a ton of note samples of all different types but consisted of one per diagnosis/disease², providing a wide sample for developing a background healthcare model, but does still not solve our data needs for tasks 4-6. A dataset within the single context of “gout” was found but further investigation revealed that this data was only a short free-text sentence of two of the chief complaint rather than a full note. This leaves me with using the publicly available datasets for the background model tasks and I will use a private dataset of history and physical examinations of patients diagnosed with severe sepsis to help build out the additional functionality and testing. This dataset will not be available in the GitHub, but the results will be.

As numbered, tasks 2-6 are still pending while I consolidate a private dataset as described and in the meantime I can continue work on tasks 2 and 3. As stated previously, the main issue I am running into is finding a focused set of healthcare notes of the same type, with the same diagnosis, de-identified, and publicly available so I have decided to drop the last requirement to facilitate my on-going work.

References

1. <https://www.med.unc.edu/medclerk/education/eduactivities/h-p-examples/>
2. <https://mtsamples.com/site/pages/browse.asp?type=89-Discharge%20Summary>