Improving STS Risk Assessment through Text-Mining of Electronic Health Records
Jacob Albert (for BS in Computer Science, Class of 2018)

Abstract:
In this project, I propose to work with Dr. Wade Schulz at the Yale School of Medicine to extract information from patients’ electronic health records relevant to assessing their risk of surgery. We will use various NLP techniques and work off of previous research to develop semantic evaluations of clinical records, particularly in devising software that establishes temporal relations between events in a patient’s medical history. If time allows, we also hope to identify and extract information from patient records that improves the risk assessment performance of the current, widely-used model developed by the Society of Thoracic Surgeons.

Background and Proposed Work:
STS developed an online tool to evaluate the risk of patients who are to undergo various surgeries (Aortic Valve Replacement, Mitral Valve Replacement, with and without Coronary Artery Bypass, etc.). Risk is assessed on a number of factors, including basic information such as sex and age, as well as information available only through patients’ electronic health records, such as events that occurred during previous operations and their temporal relationship (for instance, how much overlap there was between two events).
Currently, these data are input manually by nurses reading through the EHRs, which are unstructured English-language notes recorded by multiple doctors over the course of a patient’s life. We want to automate the process, working off of previous work in text-mining of clinical records done through a variety of techniques, including traditional NLP techniques such as Hidden Markov Models, as well as neural-network based approaches.

An additional goal in Dr. Schulz’s lab is to identify better risk assessment factors, and develop NLP extractions of patient records on these factors to improve assessments beyond the current STS model. Whether or not the extracted information is identical to that required for the STS site, the goal is to extract some meaningfully predictive information from the unstructured patient notes.

A promising tool to begin looking at is CLAMP (Clinical Language Annotation Modelling and Processing Toolkit), software based on work done by researchers at UTHealth, whose team won that Temporal Evaluation task at the 2016 International Workshop on Semantic Evaluation. Another tool that may be of use is Apache cTakes (clinical Text Analysis and Knowledge Extraction System), which also offers a competitive temporal evaluations module.

Data Set Description:

Dr. Schulz and his team have access to all clinical data, including pathological reports and unstructured notes on patients, from Yale New Haven Hospital.

References:
Lee, Hee-Jin et al. UTHealth at SemEval-2016 Task 12: an End-to-End System for Temporal Information Extraction from Clinical Notes. Link

SemEval-2016 Task 12: Clinical TempEval. Link