Identifying Patients with Pneumonia from Text for Cohort Identification

Domain Area: Evidence-based Medicine

DBMI Course: Biomedical Text Processing: Patterson, Hurdle

Problem context: Both opioids and benzodiazepines plausibly could increase risk of community-acquired pneumonia possibly secondary to sedation or aspiration. Pneumonia is a common infection with serious consequences in older adults. A healthcare delivery system has asked us to determine whether use of opioids is associated with increased risk of community-acquired pneumonia in older adults in their population. Manual chart review on tens of thousands of patients is infeasible, so we will apply natural language processing to decrease the chart review burden and design a study to compare pneumonia prevalence in patients with and without opioids and benzodiazepines.¹

Domain Learning Objective: Apply a case definition of pneumonia to determine whether elderly patients with a history of opioid or benzodiazepine use are more likely to have pneumonia than patients without a history of pneumonia.

Data Science Learning Objectives:

1. Ask Questions: Determine what clinical evidence is sufficient to label a patient as potentially having pneumonia.
2. Acquire and Assimilate Data
   a. Design a cohort selection algorithm using ICD codes.
   b. Apply text wrangling techniques to split reports into sections, sentences, and tokens.
3. Analyze Data and Answer Questions:
   a. Apply regular expressions to identify potential pneumonia cases
   b. Apply machine learning algorithms to classify reports based on pneumonia
   c. Design a study to compare pneumonia prevalence in patients with and without exposure to opioids/benzodiazepines
4. Assess results and Advise:
   a. Assess feasibility of using the NLP tool as a surrogate or support for manual chart review
   b. Design method to present NLP results to domain experts during chart review

Data sets: Chest radiograph reports from MIMIC II deceased patient set (publicly available).

Data science resources: (1) Jupyterhub notebooks with Python packages, (2) pyConText NLP algorithm.