CS 410
Text Information Systems

Final Project Presentation

Hanyin Wang , Yuanpei Ma, Yeyu Ma

Apply State-of-the-art Text Retrieval Methods on Electronic Medical Records

Background

- Electronic Medical Records (EMR) represents extremely rich text data generated from morden healthcare process.

- Faces unique challenges in text retrieval including:
  - Long documentation.
  - Unstructured data with repetitive mentioning of terms.
    - E.g., in a discharge summary “coronary artery disease” could be mentioned under both “Past Medical History” and “Current Hospital Treatment”, where only the latter matters in text retrieval process.
  - Frequent use of abbreviations.
    - “Vanco” short for “vancomycin”, “a-fib” short for “atrial fibrillation”.

Project Overview

- We hope to evaluation the effectiveness of state-of-the-art text retrieval method on EMRs.

- Testing scenario:
  - Given a collection of hospital discharge summaries, identify candidate patients with certain diagnosis and treatment for enrollment into future clinical trials.
  - E.g., with the query of “pneumonia ceftraxone”, we wish to identify all patients treated with ceftiaxone for pneumonia during the CURRENT hospital stay.
  - Should not include diagnosis from past medical history but not relevant in current hospitalization.

Database and relevance judgement

- We obtained a collection of discharge summaries from MIMIC-III database.
  - MIMIC-III ('Medical Information Mart for Intensive Care') is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital.
  - https://physionet.org/content/mimiciv/0.4/
  - This is a restricted-access resource and researcher has to complete CITI Data or Specimens Only Research training to access.
  - As the data cannot be shared, we could only run the code locally, however will provide source code to be run on any other text data collections.

- A licensed Internal Medicine Physician reviewed total 100 hospital discharges summaries on 5 queries to judge relevance.

A sample discharge summary

Past Medical History:
COPD flare [**6-7**] s/p intubation, s/p distal tracheal to Left Main
Stem stents placed [**2118-6-9**]. Stents d/c'd [**2119-4-19**], CAD w/ atypical
angina (LAD 30%, RCA 30%, EF 63%), ^chol, hypothyroidism, htn,
hiatal hernia, lacunar CVA, s/p ped struck -> head injury & rib
fx, depression
PMH:
COPD, s/p admit [**6-7**] for exacerbation requiring intubation
tracheobronchomalacia, s/p bronchial stenting
Large hiatal hernia
Lacunar CVA
Hypothyroidism by records in CCC, although patient denies and is not taking any
medication
Depression
MVA, s/p head injury approximately 10 years ago
Hypertension
Hysterectomy

Social History:
Social History: The patient is married and worked as a clinical
psychologist. Her husband is a pediatric neurologist at
[**Hospital3 **]. They have several children, one of which is a nurse.

Family History:
Family History: (+) FHx CAD; Father with an MI in his 40's, died of a CVA at age 59

Physical Exam:
Admit H+P
General-lovely 81 yr old feamle in NAD.
Neuro- intermittently anxious, MAE, PERRLA, L eye ptosis,
symetrical smile, gossly intact.
HEENT-PERRLA, sclera anicteric, pharynx- no exud or erythema
Resp-clear upper, diffuse ronchi, intermit exp wheezes
Cor- RRR, No M, R, G
Abd- soft, NT, ND, no masses. Slight protrusion at area of
hiatal hernia
Ext- no edema or clubbing

Brief Hospital Course:
82 y/o female admitted [**2119-5-4**] for consideration of tracheoplasty. Bronchoscopy done [**5-4**] confirming severe TBM. Underwent tracheoplasty [**5-5**], complicated by resp failure d/t mucous plugging, hypoxia requiring re-intubation resulting in prolonged ICU and hospital course. Also developed right upper extrem DVT from mid line.

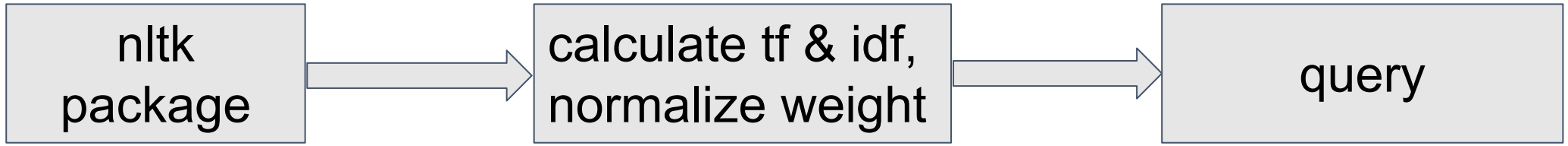
Pain- Epidural accidently d/c'd POD#1, pt briefly used dilaudid PCA intermittently w/ fair pain control. Pt required re-intubation for resp failure d/t secretions and PCA d/c at that time. Propofol for sedation while intubated. Sedation d/c'd [**5-12**] for weaning trial w/ ETT- failed trial. Trach [**5-13**]-weaning efforts as below. Minimal c/o pain since [**5-13**]. Presently pain free.

Neuro- Initially intact- post op aggitation, inhibiting weaning efforts [**5-16**]. Psych eval [**5-18**]-Started on zyprexa and ativan w/ improvement in anxiety. Presently A+Ox3- cooperative and lovely.

Resp- Extubated POD#2 then required re-intub [**5-7**] for hypoxia d/t poor cough and mucous plugging. SIMV/PS alt w/CMV at night x4-5d, with CPAP attempts during day. Bronchoscopy qd [**Date range (1) 1813**] for secretion management. Bronch [**5-9**] revealed swollen epiglottis, bronch [**5-10**] - good leak w/ ETT cuff deflated. Bronch [**5-13**] for eval/trach placement. Last bronch [**5-19**] w/ min secretions present, sputum sent.

Basic code Walkthrough

Code Walkthrough：improved VSM model



Demo

NLTK performance

Query	True Relevant Doc N	Precision @ 5	Recall @ 5	Average Precision @ 5
MRSA vancomycin	2	0	0	0
Seizure keppra	4	40%	50%	50%
AKI dialysis	2	20%	50%	16.6%
Pneumonia ceftriaxone	10	80%	40%	80%
Atrial fibrillation amiodarone	8	80%	50%	71%
Mean Average Precisions	43.5%			

Analysis of results

- NLTK model performed reasonably well in common conditions such as pneumonia and ceftriaxone, but poorly in less frequent conditions such as MRSA.

- Upon deep dive of the relevant documents not retrieved by the model, we found several areas for improvement:
  - Mismatch between query term and its abbreviation: Vanco is used instead of vancomycin.
  - Mismatch between brand name and generic name for medication: Levetiracetam vs Keppra.
  - Inclusiveness of medical terms: dialysis could include HD and ultrafiltration.

- Therefore we worked on improving search results by heuristic query expansion.

NLTK performance after query expansion

Old Query	Expanded Query	True Relevant Doc N	Precision @ 5	Recall @ 5	Average Precision @ 5
MRSA vancomycin	vanco vancomycin MRSA	2	40%	100%	32.5%
Seizure keppra	keppra Levetiracetam Seizure	4	60%	75%	60.4%
AKI dialysis	ultrafiltration HD AKI	2	40%	100%	100%
Pneumonia ceftriaxone	Pneumonia ceftriaxone	10	80%	40%	80%
Atrial fibrillation amiodarone	Atrial fibrillation amiodarone	8	80%	50%	71%
Mean Average Precisions		69.4%			

Conclusions

- Text retrieval of EMRs require special considerations out of traditional BM25 models.

- Areas of improvement for future model could include:
  - Query expansion to include common abbreviations, generic names etc.
  - Query substitution to replace general terms to more specific terms: e.g., HD, ultrafiltration for dialysis, cholecystectomy for surgery.

Thank You For Watching!

f

The Grainger College of Engineering

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN