PAPER ID 137

AUTOMATED ICD-9-CM MEDICAL CODING OF DIABETIC PATIENT'S CLINICAL [1)]

{kumar64, nmistry2}@illinois.edu

CS 598: Deep Learning for Health, Spring 2022
University of Illinois, Urbana-Champaign

1) Vitor Pereira, Sérgio Matos, and José Luís Oliveira. 2018. Automated ICD-9-CM medical coding of diabetic patient's clinical reports. In International Conference on Data Science, E-learning and Information Systems 2018 (DATA '18), October 1–2, 2018, Madrid, Spain. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3279996.3280019

Problem Statement

- All patient data are stored as Electronic Health Record (EHR), which is important for better patient care – diagnosis and treatment and contains health data across time
- EHR is used by other health industry players such as insurance for cost reimbursement
- Codes are used in EHR to document various diagnosis and procedures used for patient care and treatment. ICD-9-CM and ICD-10-CM are industry standard coding.
- There are cases where ICD codes needs to be entered for missing codes or corrections needs to be done, based on notes in EHR records
- Manual entry of ICD codes is a cost intensive process and needs to be done by experts. In US, it is estimated to cost approx. 25 Billion USD per year.

Proposal

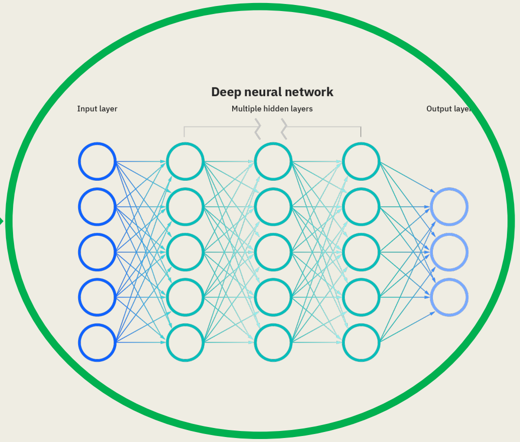
- The author of the paper «Automated ICD-9-CM Medical Coding of Diabetic Patient's Clinical» proposes to deploy deep learning models to identify ICD-9-CM codes from the EHR notes using multi-label classification
- The solution complements the current process and help health industry achieve higher quality data in EHR together with cost reduction

Cancer (Malignant Neoplasm), Hepatic (Liver) Assessment: Patient is more lethargic yesterday & today than he was on Fri ([**2-10**] days ago) Action: He was made DNR/CMO tonight, per agreement of family.

Assessment: Patient had acute SOB, midsternal chest pain, feeling that he was going to die Q **2016**] when he rolled in bed onto bedpan & had Br. HR increased to lon 70s SR. BP increased to 149/systolic. Desatted to 85%

Action: Given 100% high flow neb, 0.5 NTP & 0.25mg IV morphine EKG done during SOB. Response: Pain & SOB relieved. No changes on EKG

Plan: Now that patient is CMO, medicate W/morpnan rolling patient in bed. Continue to medicate w/Lopressor to prevent ACS as well as NTP or SL NTG, morphine



Deep learning model for generating ICD codes from EHR Notes

{ 173.80, 173.89, 173.90 }

ICD-9-CM Code

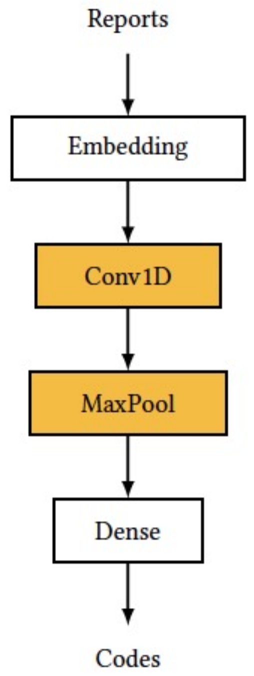
Training Dataset

- MIMIC III
  - DIAGNOSIS_ICD and NOTEEVENTS tables
- Preprocessing
  - Data for patients with diabetes
  - ICD-9-CM codes – regular and rolled up version
  - Tokenization of notes
    - Punctuation characters removal
    - Digits replacement with ‘d’
    - Lowercase
    - Tokenization at whitespace
  - Removal of all notes with less than 10 tokens and more than 2199 tokens
  - Replacement of infrequent word with frequent word with closest Levenstein distance

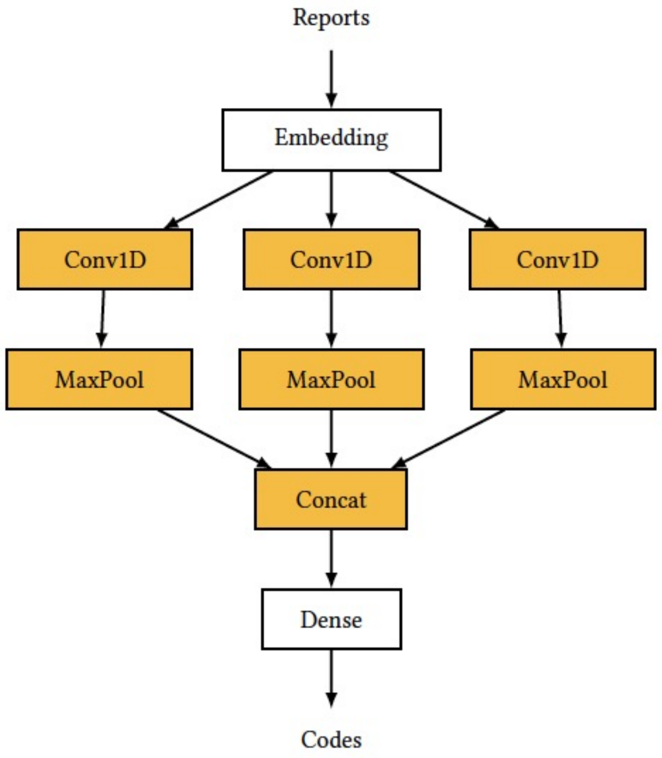
Num. of used records	399623
Num. of regular labels	4103
Num. of rolled up labels	781
Num. of unique tokens	53229
Avg. num. of tokens per report	309.05638814582744

Models

- Bag of Tricks – linear model
- Convolutional Neural Network (CNN)
  - Baseline
  - Three parallel convolutional layer
- Implementation is done using keras library



(a) Baseline



(b) Three parallel convolutional layers

Claims and results

1. CNN Baseline has higher Precision, Recall and F1 as compared to BoT Baseline
2. CNN 3-Conv1D has higher Precision, Recall and F1 than CNN Baseline
3. Precision, Recall and F1 for Rolled up ICD-9-CM code assignment is always higher than Regular ICD-9-CM code assignment

Result 1

	Reprod.	Original
BoT BaseLine Regular (Precision, Recall, F1)	63.79, 5.04, 9.34	66.25, 8.61, 15.24
CNN Baseline Regular (Precision, Recall, F1)	72.03, 15.82, 25.94	73.97, 25.88, 38.13

Result 2

	Reprod.	Original
CNN Baseline Regular (Precision, Recall, F1)	72.03, 15.82, 25.94	73.97, 25.88, 38.13
CNN 3-Conv1D Regular (Precision, Recall, F1)	75.82, 30.71, 43.71	76.07, 31.46, 44.51

Result 3

	Reprod.	Original
BoT Baseline Regular (Precision, Recall, F1)	63.79, 5.04, 9.34	66.25, 8.61, 15.24
BoT Baseline Rolled (Precision, Recall, F1)	85.67, 10.32, 18.43	75.91, 18.89, 30.25
CNN Baseline Regular (Precision, Recall, F1)	72.03, 15.82, 25.94	73.97, 25.88, 38.13
CNN Baseline Rolled (Precision, Recall, F1)	77.51, 28.68, 41.87	77.73, 35.13, 48.38
CNN 3-Conv1D Regular (Precision, Recall, F1)	75.82, 30.71, 43.71	76.07, 31.46, 44.51
CNN 3-Conv1D Rolled (Precision, Recall, F1)	79.56, 38.33, 51.73	79.82, 38.26, 51.73

Ablation

- Trainable initial weight in the first layer of CNN models – better result as compared to fix initial weight

	Fixed Initial Weight.	Trainable Weight
CNN Baseline Regular (Precision, Recall, F1)	72.03, 15.82, 25.94	74.98, 21.91, 33.91
CNN Baseline Rolled (Precision, Recall, F1)	77.51, 28.68, 41.87	79.14, 44.66, 57.1
CNN 3-Conv1D Regular (Precision, Recall, F1)	75.82, 30.71, 43.71	75.95, 32.44, 45.46
CNN 3-Conv1D Rolled (Precision, Recall, F1)	79.56, 38.33, 51.73	81.13, 43.56, 56.68

Challenges

- Pre-processing of MIMIC III data: Time consumed during pre-processing specially in replacing infrequent tokens with frequent ones using Levenstein distance and Word2Vec model training
- Training of models: Computation resource restriction on personal notepad and on Google Colab Pro led us to adapt improvised training approach – training with 20% of training data at a time and then saving the intermediate model for next training

Thank you

Although we faced several challenges during the project, it was great learning experience, which required us to think out-of-box to resolve computational resource restriction.

Thank you for the opportunity to work on a research paper experiment and establish the paper baseline.