Readmission prediction via deep contextual embedding of clinical concepts

Gowtham Kuntumalla (gowtham4)

Yiming Li (yl159)

Course Project
UIUC CS598 - Spring 2023

https://doi.org/10.1371/journal.pone.0195024

Outline

- Introduction to the paper
- Description of data
- Deep Learning Model
- Results
  - Comparison study
- Demo of working code

Introduction to the paper

17.6% patients readmitted < 30D after discharge from hospital

76% cases avoidable

$17.9 billion/year wasted medicare spending

US Healthcare: http://doi.acm.org/10.1145/2783258.2788585

Data Description - Electronic Health Records (EHR)

```
PID DAY_ID  DX_GROUP_DESCRIPTION    SERVICE_LOCATION    OP_DATE
1   73888   ANGINA PECTORIS DOCTORS OFFICE  74084
1   73888   MONONEURITIS OF UPPER LIMB AND MONONEURITIS MULTIPLEX   DOCTORS OFFICE  74084
1   73888   SYMPTOMS INVOLVING RESPIRATORY SYSTEM AND OTHER CHEST SYMPTOMS  DOCTORS OFFICE  74084
1   73880   ACUTE APPENDICITIS  INPATIENT HOSPITAL  74084
1   73880   DIABETES MELLITUS   INPATIENT HOSPITAL  74084
1   73880   ESSENTIAL HYPERTENSION  INPATIENT HOSPITAL  74084
1   73880   OTHER FORMS OF CHRONIC ISCHEMIC HEART DISEASE   INPATIENT HOSPITAL  74084
1   74450   EZETIMIBE   Pharmacy_Claim  74084
1   74498   CATARACT    DOCTORS OFFICE  74084
1   74498   OTHER AND ILL-DEFINED CEREBROVASCULAR DISEASE   DOCTORS OFFICE  74084
1   74498   VISUAL DISTURBANCES DOCTORS OFFICE  74084
1   73999   ESSENTIAL HYPERTENSION  DOCTORS OFFICE  74084
1   73999   OTHER AND UNSPECIFIED DISORDERS OF BACK DOCTORS OFFICE  74084
1   73999   OTHER FORMS OF CHRONIC ISCHEMIC HEART DISEASE   DOCTORS OFFICE  74084
1   73999   OTHER PERSONS SEEKING CONSULTATION WITHOUT COMPLAINT OR SICKNESS    DOCTORS OFFICE  74084
1   74517   METOPROLOL TARTRATE Pharmacy_Claim  74084
1   74328   POTASSIUM CHLORIDE  Pharmacy_Claim  74084
1   73657   DISORDERS OF MUSCLE, LIGAMENT, AND FASCIA   DOCTORS OFFICE  74084
1   74084   ANGINA PECTORIS DOCTORS OFFICE  74084
1   74084   HEART FAILURE   DOCTORS OFFICE  74084
1   74084   NONSPECIFIC ABNORMAL RESULTS OF FUNCTION STUDIES    DOCTORS OFFICE  74084
1   74084   SYMPTOMS INVOLVING RESPIRATORY SYSTEM AND OTHER CHEST SYMPTOMS  DOCTORS OFFICE  74084
1   74084   ANGINA PECTORIS OUTPATIENT HOSPITAL 74084
1   74084   HEART FAILURE   OUTPATIENT HOSPITAL 74084
1   74084   NONSPECIFIC ABNORMAL RESULTS OF FUNCTION STUDIES    OUTPATIENT HOSPITAL 74084
1   74084   SYMPTOMS INVOLVING RESPIRATORY SYSTEM AND OTHER CHEST SYMPTOMS  OUTPATIENT HOSPITAL 74084
```

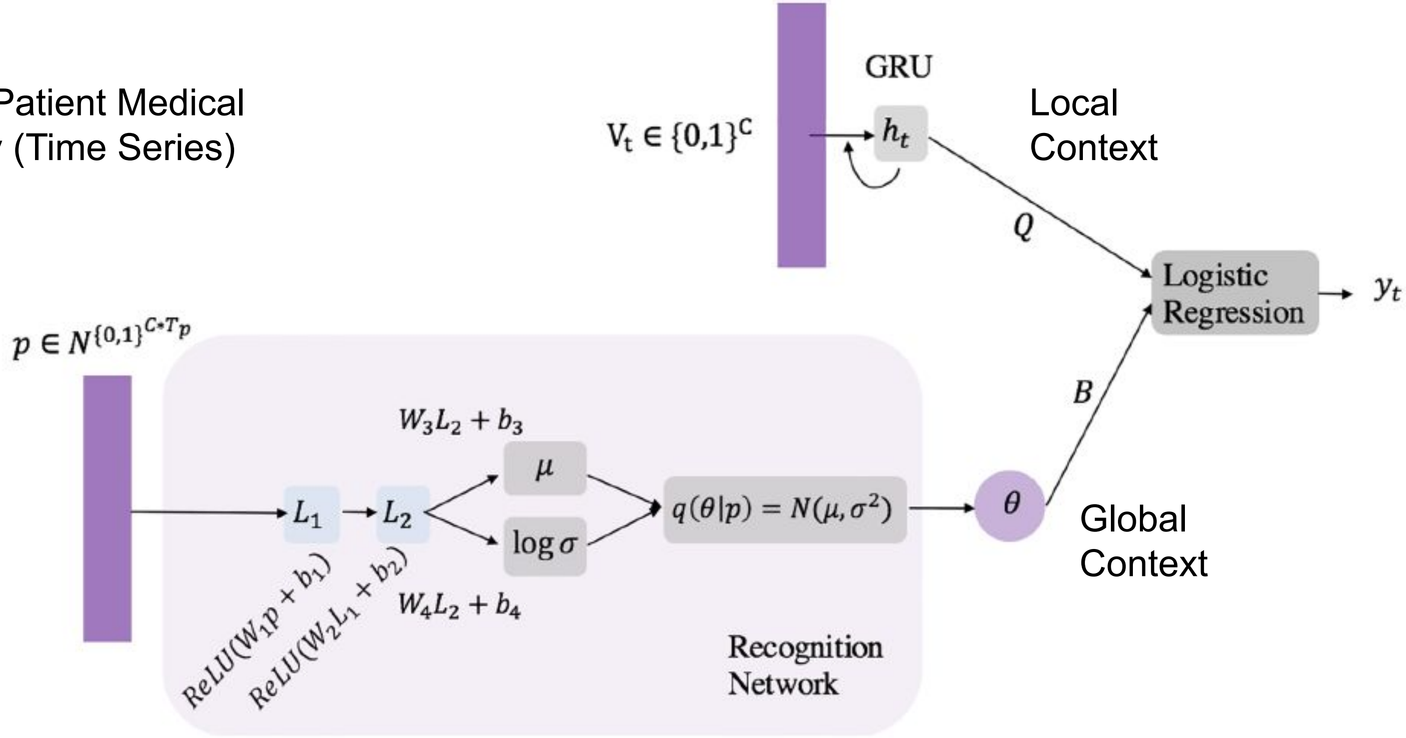
- Access to synthetic dataset only
- 3000 Patients, Max 300 Visits/Patient
- ~ 600 Distinct medical concepts

Columns

- Patient ID (patient)
- Day ID (visit)
- Occurrence Info (Event) *aka* medical concepts
- Location

CONTENT Model = GRU + TopicRNN (RecognitionNet)

Input: Patient Medical History (Time Series)



Local Context

Binary Classification Task

Global Context

What does Topic Modeling mean?

Table 1: Five Topics from the TopicRNN Model with 100 Neurons and 50 Topics on the PTB Data. (The word *s&p* below shows as *sp* in the data.)

Law	Company	Parties	Trading	Cars
law	spending	democratic	stock	gm
lawyers	sales	republicans	s&p	auto
judge	advertising	gop	price	ford
rights	employees	republican	investor	jaguar
attorney	state	senate	standard	car
court	taxes	oakland	chairman	cars
general	fiscal	highway	investors	headquarters
common	appropriation	democrats	retirement	british
mr	budget	bill	holders	executives
insurance	ad	district	merrill	model

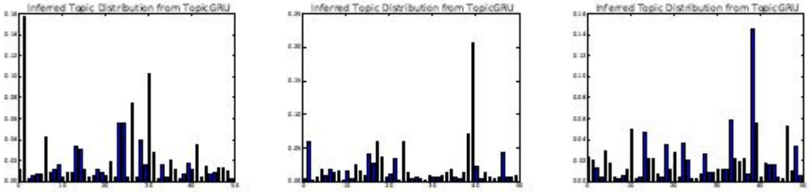


Figure 2: Inferred distributions using TopicGRU on three different documents. The content of these documents is added on the appendix. This shows that some of the topics are being picked up depending on the input document.

- For Semantic Coherence!
- Helps give global context to the patient medical health records so far
- Via Latent Topics
- Thereby improves time series predictions

Taken from paper- ”TOPICRNN: A RECURRENT NEURAL NETWORK WITH LONG-RANGE SEMANTIC DEPENDENCY”

https://doi.org/10.48550/arXiv.1611.01702

Results - Comparison Study

Table 2. Performance comparison on CHF data. CONTENT outperforms Word2vec+LR, Med2vec+LR, GRU, GRU+Word2Vec, and RETAIN on different performance metrics.

Method	PR-AUC	ROC-AUC	ACC
Word2vec+LR	0.3445±0.0204	0.5360±0.0246	0.6828±0.0120
Med2vec+LR	0.3836±0.0149	0.5937±0.0120	0.6915±0.0095
GRU	0.3862±0.0136	0.5998±0.0124	0.6856±0.0082
GRU+Word2Vec	0.3430±0.0157	0.5616±0.0157	0.6731±0.0091
RETAIN	0.3720±0.0148	0.5707±0.0140	0.6814±0.0111
CONTENT	0.3894±0.0153	0.6103±0.0130	0.6934±0.0090

Table 3. Performance comparison on synthetic data. CONTENT outperforms Word2vec+LR, Med2vec+LR, GRU, GRU+Word2Vec, and RETAIN on different performance metrics.

Method	PR-AUC	ROC-AUC	ACC
Word2vec+LR	0.5155±0.0021	0.6040±0.0188	0.6229±0.0179
Med2vec+LR	0.5906±0.0057	0.6884±0.0044	0.7170±0.0087
GRU	0.5929±0.0100	0.6881±0.0048	0.7141±0.0040
GRU+Word2Vec	0.5907±0.0174	0.6836±0.0031	0.7117±0.0045
RETAIN	0.5525±0.0005	0.6927±0.0001	0.7310±0.0001
CONTENT	0.6011±0.0191	0.6886±0.0074	0.7170±0.0069

https://doi.org/10.1371/journal.pone.0195024.t003

Results:
K(20)-Means Clustering using representations

Patient-specific vector representation
=
Topical context vector θ
+
final hidden state of RNN

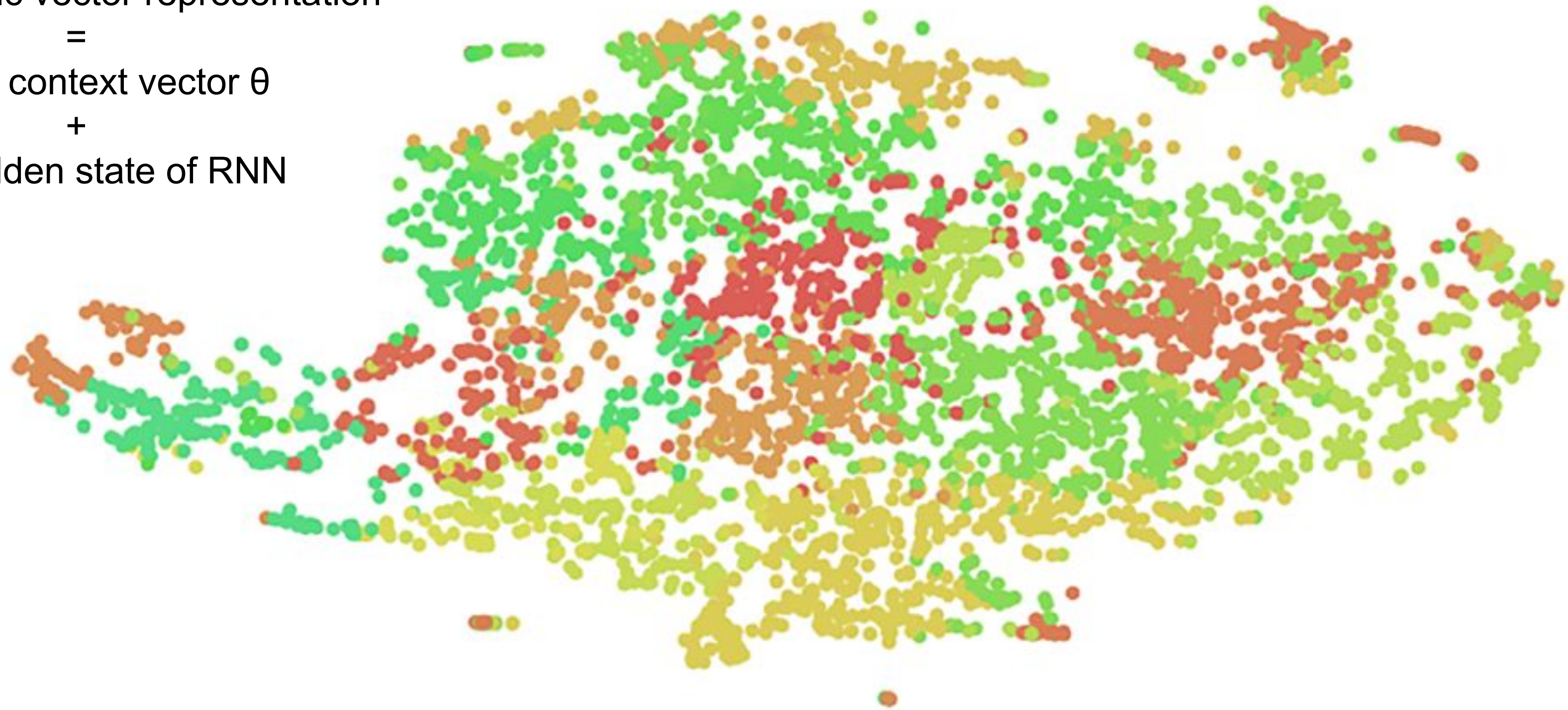
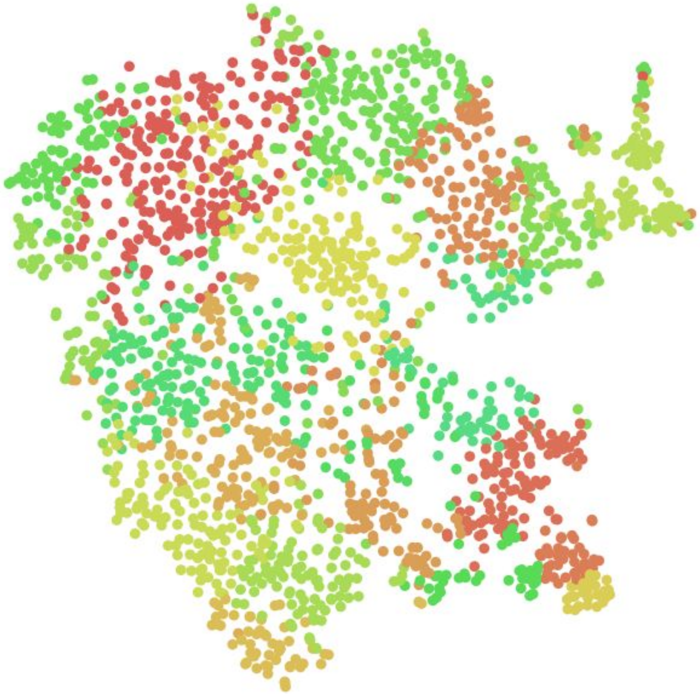


Fig 3. Clustering of patient representations.

Another TSNE cluster



Demo

In the video, we talk about our work on replicating the results from this paper and additional experimentation.

Youtube Link: https://youtu.be/tdRc92b4_EI

References:

1. https://doi.org/10.1371/journal.pone.0195024
2. https://doi.org/10.48550/arXiv.1611.01702
3. Original authors Github Repo: https://github.com/danicaxiao/CONTENT
4. Our code reproduction: https://github.com/yl159-Yiming/CS598-DL4H