Readmission prediction via deep contextual embedding of clinical concepts

Ashwani Singh and Mohammad Hassanpour

{ashwani3, mh57}@illinois.edu

Group ID: 68

Paper ID: 168

Presentation link: https://www.youtube.com-Workinprogress

Code link: https://www.github.com-Workinprogress

1

Introduction

3 Methodology

We reproduce a paper ([Xiao et al., 2018](#br3)) on read-

mission prediction via deep contextual embedding

of clinical concepts. Being discharged from a hospi-

tal and being admitted within thirty days indicates

poor quality care, is synonymous with poor patient

health, and is rather costly for the payer, be that a

health plan, the patient, or a state or federal gov-

ernment. Accurate prediction allows providers to

target patients at greater risk and reduce readmis-

sion.

3

.1 Model descriptions

CONTENT is a hybrid deep learning model that

predicts hospital readmission by transforming com-

plicated patient event structures from the EHR

records as deep clinical concept embedding. These

encodings represents both long and short term con-

ditions of the patient, which can be viewed as

global and local context.

CONTENT is an extension of the contextual RNN model. It consists of a Gated Recurrent Unit (GRU) and a topic model as a recognition network. The RNN model computes the hidden states for each patient visit (Vt) at time “t”. Each of these contains a subset of medical events. These subsets of events (i.e. short term dependencies) during a patients visit is considered the “local context” and the goal of the CONTENT model is to help RNN (GRU) perform better in the local context.

The paper ([Xiao et al.](#br3), [2018)](#br3) trains a TopicRNN

model on Electronic Health Records (EHR) related

to Congestive Heart Failure (CHF) for this predic- The RNN model computes the hidden states for

tion. RNNs and latent topic models are combined in the TopicRNN model to capture local (syntactic) and global (semantic) dependencies. RNNs are good at capturing the local structure of a word sequence, but they may have difficulty remembering long-range interactions. Word order is not considered in latent topic models, which capture the general structure of a document.

2

Scope of reproducibility

2

.1 Addressed claims from the original paper

•

The CONTENT model outperforms the base- patient subgroups with more homogeneous latent

line models.

patterns. Metrics and measures such as

the average number of re-admissions and top clinical events can be calculated for each subgroup to help determine the causes of readmissions.

•

By leveraging topics to capture global context,

the CONTENT model has the capability to

alleviate the effects of missing values.

“The CONTENT model proposed here and

trained on Congestive Heart Failure (CHF)

Electronic Health Record (EHR) data out-

performs Word2vec+LR, Med2vec+LR, GRU,

GRU+Word2Vec, and RETAIN on PR-AUC, ROC-

AUC, and ACC performance metrics.”

This is achieved with the help of the recognition

network (a multi-layer perceptron), which provides

a context vector (or global context) representing

a patient’s distributed medical history. The global

context is used as additional bias to the output of

the RNN model, so that the RNN can focus its

modeling capacity on the local context and provide

more accurate prediction.

Once we calculate the hidden states for each patient and visit, we concatenate this context vec-

tor θ and the ﬁnal hidden state of RNN as the

patient-speciﬁc vector representation. These vec-

tors are then used to cluster the CHF cohort into

3

.2 Data descriptions

to match the same accuracy achieved by the authors

using the synthetic data.

The authors conducted the experiment using real

world EHR data of Congestive Heart Failure (CHF)

cohort including 5,393 patients. The data included

disease, lab test, and medication codes, all binary

encoded indicating their presence or absence. The

CHF cohort was constructed according to the fol-

lowing criteria 1) ICD-9 diagnosis of heart failure

appeared in the EHR for at least two outpatient

encounters, indicating consistency in clinical as-

sessment, and 2) At least one medication was pre-

scribed with an associated ICD-9 diagnosis of heart

failure. In addition, the diagnosis date was deﬁned

as its ﬁrst appearance in the record.

3

.5 Computational requirements

The authors of CONTENT did not provide infor-

mation on the computational requirements.

We are planning to use Google Colaboratory’s

free tier GPU - ”Tesla T4” with 16GB VRAM

and running NVIDIA (R) Cuda compiler (Cuda

compilation tools) release 11.8, V11.8.89. We are

yet to clock the time taken for training per epoch.

We are estimating it to be close to 60 secs per epoch.

Additionally, we will be calculating the CPU/GPU

time taken and memory usage once we are ready

with the code.

In addition to the CHF data, the authors evalu-

ated the model on a synthetic EHR data simulated

from a de-identiﬁed real world patient dataset. The

authors generated the synthetic data as follows: For

each original patient record they randomly sample

4

Results

The researchers of the CONTENT model compared

the performance of the CONTENT model with

baseline models like Word2Vec, Med2Vec, GRU,

GRU+Word2Vec and RETAIN, and the proposed

CONTENT model outperformed the baseline mod-

els on all performance metrics. Since the model

is a binary classiﬁcation, the authors captured fol-

lowing evaluation metrics: PR-AUC, ROC-AUC

and Accuracy of the proposed model and other

baselines.

3

0% to 50% of the visits in that record and drop

the un-sampled visits. After subsampling, they per-

mute patient index. Next, for each new patient

record, they randomly combined it with another

new record, with the event time of the second pa-

tient record being aligned to the ﬁrst one.

Following this approach, the authors generated

3

000 synthetic patients, of which 2000 are used

in model training, 500 for validation, and 500 for

testing

In absence of the real world EHR data, We

will be using the Synthetic data for evaluating the

model.

In addition, the authors showcased that the CON-

TENT model has the ability to reduce the impact

of missing values by leveraging topics to capture

global context.

4

.1 Result 1

Our aim is to validate the paper’s ([Xiao et al., 2018](#br3))

TODO” - Describe how you set the hyperparam- conclusion by attaining an accuracy level that falls

3

.3 Hyperparameters

“

eters and what the source was for their value (e.g.

paper, code or your guess).

within 1% of the reported value. This indicates that

incorporating the global context during the training

of the local context led to an improvement in the

prediction’s performance or outcome

3

.4 Implementation

The code provided in the paper ([Xiao et al., 2018](#br3))

is outdated. For example ”MaskingLayer” and

Note: We were unable to execute the code pro-

vided in the paper ([Xiao et al.](#br3), [2018](#br3)) as it needs

”

ThetaLayer” from Lasagne are no more supported. lot of rewriting/refactoring. We plan to submit the

We will be writing our own code to reproduce the

proposed CONTENT model. A GitHub link will

be provided as part of the ﬁnal project submission.

The results published in the paper [(Xiao et al.](#br3),

results as part of the ﬁnal submission in a tabular

format, comparing the performance of each model.

4

.2 Result 2

[2](#br3)

[018](#br3)) outperforms most baselines due to it models

The CONTENT model can mitigate the impact of

missing values by utilizing topics to capture global

context. This claim is supported by evidence ob-

tained by introducing dropout and randomly drop-

ping a few visits.

patient representations and predict re-admissions

not only based on the RNN states but also on the

topics.

As part of this reproduction project, our goal is

Note: We were unable to execute the code pro- 5.2 What was difﬁcult

vided in the paper ([Xiao et al.](#br3), [2018](#br3)) as it needs

TODO - Describe which parts of your reproduction

lot of rewriting/refactoring. We plan to submit the

results as part of the ﬁnal submission in a tabular

format, comparing the performance of each model

with dropout.

study were difﬁcult or took much more time than

you expected. Perhaps the data was not available

and you couldn’t verify some experiments, or the

author’s code was broken and had to be debugged

ﬁrst. Or, perhaps some experiments just take too

much time/resources to run and you couldn’t verify

them. The purpose of this section is to indicate

to the reader which parts of the original paper are

either difﬁcult to re-use, or require a signiﬁcant

amount of work and resources to verify.

4

.3 Additional results not present in the

original paper

We are planning to

•

Introduce dropout technique to showcase the

low impact due to missing values.

Tips: Be careful to put your discussion in con-

text. For example, don’t say “the math was difﬁcult

to follow,” say “the math requires advanced knowl-

edge of calculus to follow.”

•

•

Perform few ablations

Tune few hyperparameters

5

.3 Recommendations for reproducibility

Note: We were unable to execute the code pro-

vided in the paper ([Xiao et al.](#br3), [2018](#br3)) as it needs

lot of rewriting/refactoring. We plan to provide

detailed experiments conducted and its results as

part of the ﬁnal submission.

TODO - Describe a set of recommendations to the

original authors or others who work in this area for

improving reproducibility.

6 Communication with original authors

TODO - Document the extent of (or lack of) com-

munication with the original authors. To make sure

5

Discussion

TODO - Describe larger implications of the ex- the reproducibility report is a fair assessment of the

perimental results, whether the original paper was

reproducible, and if it wasn’t, what factors made it

irreproducible.

Give your judgement on if you feel the evidence

you got from running the code supports the claims

of the paper. Discuss the strengths and weaknesses

of your approach – perhaps you didn’t have time

to run all the experiments, or perhaps you did ad-

ditional experiments that further strengthened the

claims in the paper.

original research we recommend getting in touch

with the original authors. You can ask authors spe-

ciﬁc questions, or if you don’t have any questions

you can send them the full report to get their feed-

back.

References

Cao Xiao, Tengfei Ma, Adji B Dieng, David M Blei, and

Fei Wang. 2018. Readmission prediction via deep

contextual embedding of clinical concepts. PloS one,

1

3(4):e0195024.

5

.1 What was easy

TODO - Describe which parts of your reproduction

study were easy. E.g. was it easy to run the author’s

code, or easy to re-implement their method based

on the description in the paper. The goal of this

section is to summarize to the reader which parts of

the original paper they could easily apply to their

problem.

Tips: Be careful not to give sweeping general-

izations. Something that is easy for you might be

difﬁcult to others. Put what was easy in context

and explain why it was easy (e.g. code had exten-

sive API documentation and a lot of examples that

matched experiments in papers).