Personalized Text Normalization of Clinical Notes

CS733: Team 12

Sanjeeva Reddy Dodlapati
Mohammad Shiri

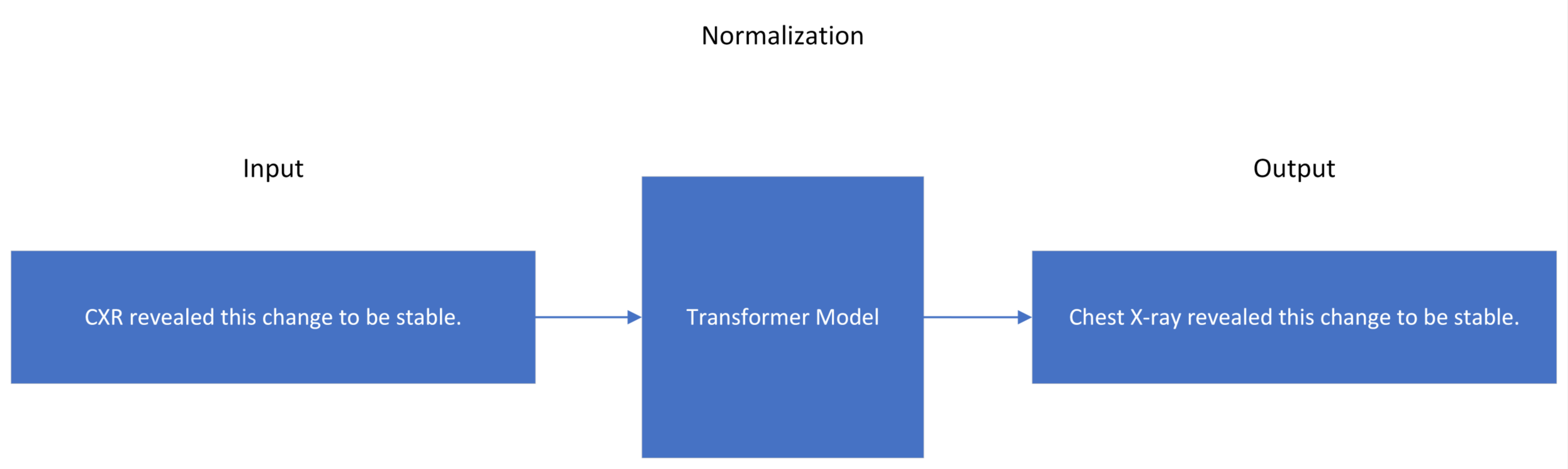
Text Normalization

Text normalization is a process of standardizing the text by replacing non-standard words or removing irrelevant components.

Before	After
A baby giraffe is 6ft tall and weighs 150 lbs	A baby giraffe is six feet tall and weighs one hundred fifty pounds
I live at 123 King Ave	I live at one two three King Ave
This looks cooooooooolllllll	This looks cool
UNDP helps reducing poverty	United Nations Development Program helps helps reducing poverty
Measurement shows 2 mA current	Measurement shows two milliamperes current

Proposed Scheme

Text normalization is a process of standardizing the text.



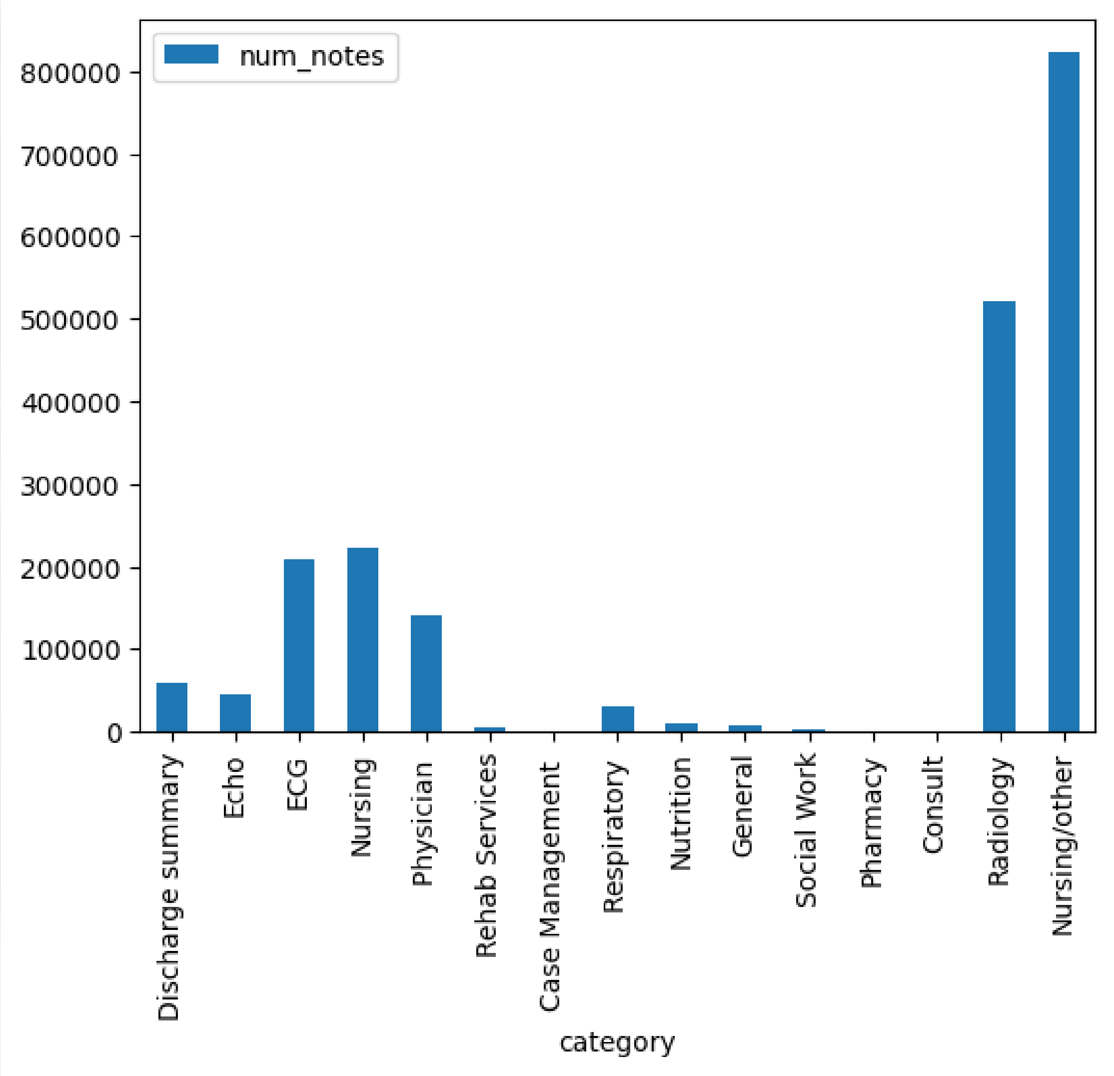
Personalized Text Normalization

- Personalized text normalization is a process where text is standardized differently for different users.
- For example, if the user is a doctor or nurse only some of the acronyms to be replace because they have substantial medical knowledge to understand them.
- On the other hand, non-medical background persons may require all the acronyms to be expanded so that they can understand their health records.

Related Work

- Deep learning-based methods to tackle text normalization
- Treated text normalization problem as a machine translation task
- Personalized text normalization has not been widely explored
- AsiaSpic: probabilistic method to support the use of user-defined short-forms in a multilingual chat system by exploiting a personalized dictionary for each user to support user-defined short-forms.

The Data: Acquisition and description



- MIMIC-III Data (used as train and validation set)
  - Clinical notes of different categories (e.g. discharge summary notes)
  - From forty thousand patients in critical care units
- n2c2(track-2) NLP Research Data (used as test set):
  - Clinical notes of medication and discharge notes
  - From multiple visits of thousands of patients
- Two dictionaries of medical terminology abbreviations

Data Preprocess

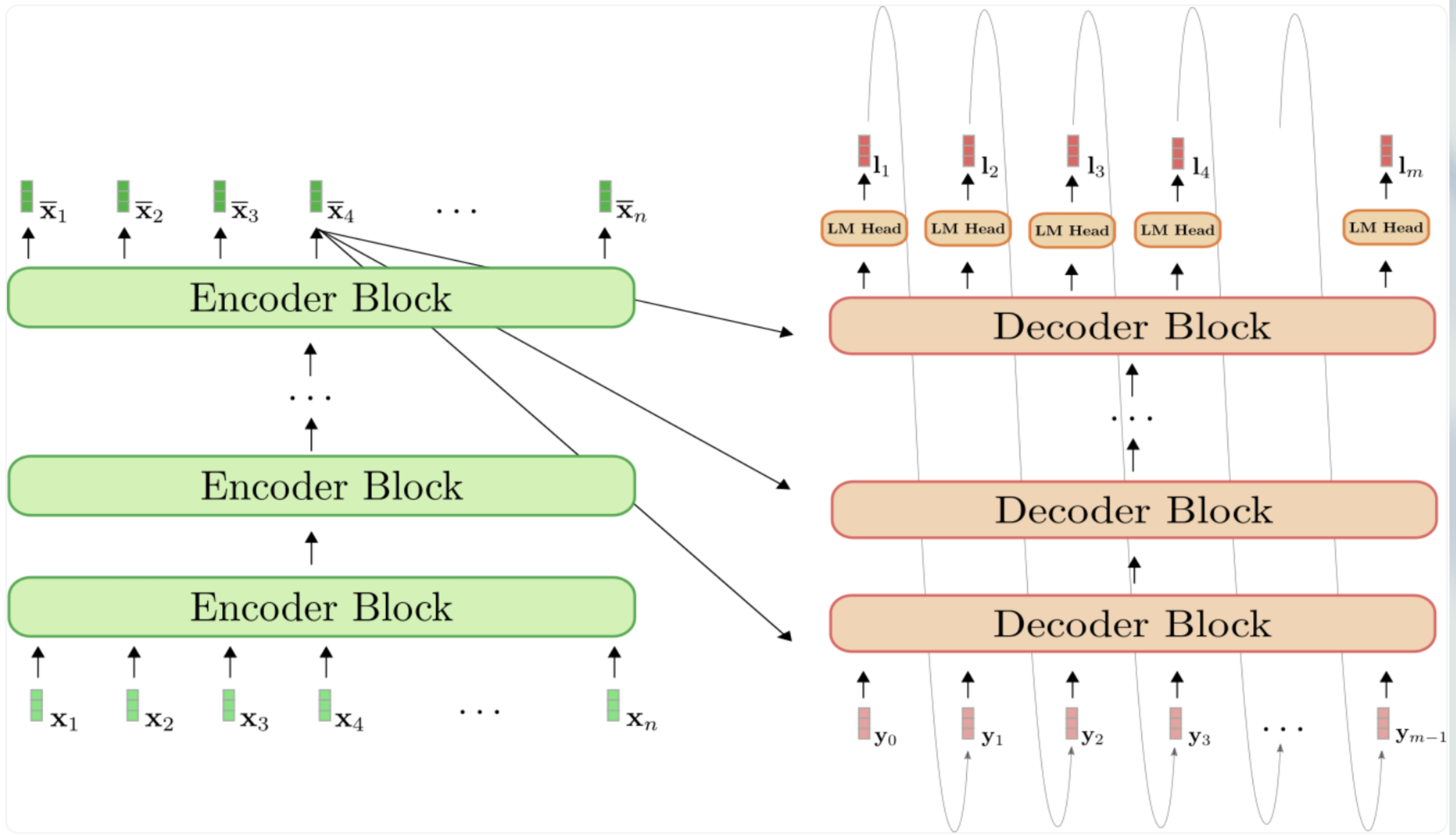
- Removed unnecessary information in text (e.g. date)
- Replaced abbreviations with their meaning from terminology dictionary(a.c. --> before meals)
- Labeled for three personalized users; non-medical, nurses, doctors
- Sequences with around 500 word long

sentences	labels
"non-medical" md dictated by: medquist. pulmicort nebulizer b.i.d. fluticasone mcg tw	"non-medical" median dictated by: medquist. pulmicort nebulizer twice daily fluticasone
"doctor" md dictated by: medquist. pulmicort nebulizer b.i.d. fluticasone mcg two puff	"doctor" md dictated by: medquist. pulmicort nebulizer b.i.d. fluticasone mcg two puffs
"nurse" md dictated by: medquist. pulmicort nebulizer b.i.d. fluticasone mcg two puffs	"nurse" md dictated by: medquist. pulmicort nebulizer twice daily fluticasone mcg two p

Screenshot of preprocessed personalized input and output examples of clinical notes

Model Architecture

- Built a Bert2bert Encoder-Decoder
- Both the encoder and decoder are initiated with the weights of ClinicalBERT, which is pretrained on MIMIC-III data
- Even though some existing studies suggest that GPT3 decoder performs better than bert decoder, we have chosen ClinicalBERT to avoid pretraining GPT3 from scratch
- Inputs are the sentences that contain abbreviations (acronyms)
- Labels are the sentences with the expanded abbreviations



Encoder-Decoder model architecture of Clinical-text-Norm-BERT

Model Training & Evaluation

- Figure on the right shows training progress on a small dataset
- ROUGE metrics are a set of metrics to calculate unigram, or bigram or multi-gram metrics
- On larger dataset (5%), maximum ROUGE-1 (unigram) precision 30%
- Recall and F1 score are hovering around 10%

[126/126 1:59:03, Epoch 2/2]

Step	Training Loss	Validation Loss	Rouge1 Precision	Rouge1 Recall	Rouge1 Fmeasure
10	6.432100	23.376423	0.043600	0.021200	0.028300
20	6.309600	24.290266	0.040600	0.019900	0.026400
30	6.202500	25.589096	0.038400	0.018800	0.025000
40	6.206100	27.121540	0.035300	0.017400	0.023100
50	6.050000	26.727184	0.034400	0.016900	0.022400
60	6.132600	28.546398	0.019100	0.009100	0.012300
70	6.028400	28.022633	0.030100	0.014200	0.019000
80	5.953800	25.116844	0.045500	0.022200	0.029600

Training metrics on small dataset

Model Training & Evaluation continued...

- Model is paying more attention to punctuation and special characters like backslash, parenthesis, star symbol
- Higher length penalty encourages model to generate longer sentences

```
Inputs ................................: ['no associated fever, no respiratory compromise.',
tory of "getting shaky" if etoh withdrawal.', 'no known seizure history.', 'given total of
idine (on this at baseline), thiamine/folate, b, magnesium (mg ), calcium (ionized ca ).',
spiratory compromise.', 'he was initially admitted to for dts, observation.', 'on arrival t
ory and physical exam.', 'given mg valium x and then mg.', 'labs recheckedica up to , repea
use with etoh and benzo use.', 'he drinks regularly and has intermittent binges of several
ad or happy.', "has been at several hospitals including for inpt detox from benzo's includi
Predictions~~~~~~~~~~~~~~~~~~~~~~~~~~~: ['the is : : to..', 'the : : to, and..', 'the was :
p.', 'thesp : : * : the, *, :,, of, -, mg, and, (, to, p,., ).', 'he was : : mg, the,, :, a
and : and, (, p, to,. and and..', 'no : :..', 'the was was : : the,, and,..', 'the patient
d and..', 'was was : : mg, :,, and, (,. and..']
```

Screenshot of input sentences to the model and predicted sentences generated by the model

Discussion

- Model performance is not up to our satisfaction. Much needs to be improved
- The probable reasons for the model unable to learn quickly:
  - Due to time and resource constrains model trained on just 5% of the data available
  - Data preprocessing was a very challenging task and further cleanup needed.
  - Predicting whole sentence instead of just abbreviation expansion.
  - Hyperparameters like learning rate and weight decay used are default setting. Needs to find optimal settings.

Future work

- Further clean up of the input data is required to remove leftover special characters
- Training model on larger dataset. May take several days to train on cluster with gpu resources.
- Normalizing all the terms including fractions, numbers, medical codes, and acronyms
- Finding optimal hyper-parameters

Conclusion

- For the first time, we have attempted to implement personalized text normalization on electronic health records.
- We have collected the clinical notes from two different databases and annotated with labels for personalized text normalization.
- We have built encoder-decoder architecture initialized from ClinicalBERT model.
- Trained the proposed model on 5% of the 2.5 million notes to achieve 30% of ROUGE-1 precision.
- Further experiments to improve model training are underway and the results will be updated in the near future.

Thank you for your kind attention!

If you have any questions or comments, we would be happy to receive them by:

mshir001@odu.edu

sdodlapa@odu.edu

Code is made available at:
https://github.com/SanjeevaRDodlapati/Clinical-Text-Norm-BERT