NLP for Healthcare

Medical Codes prediction using discharge summaries

Chua Yeow Long

Overview

- Problem Statement
- Data Preprocessing
- Model
- Baselines
- Results
- Conclusion
- Future Work

Problem Statement

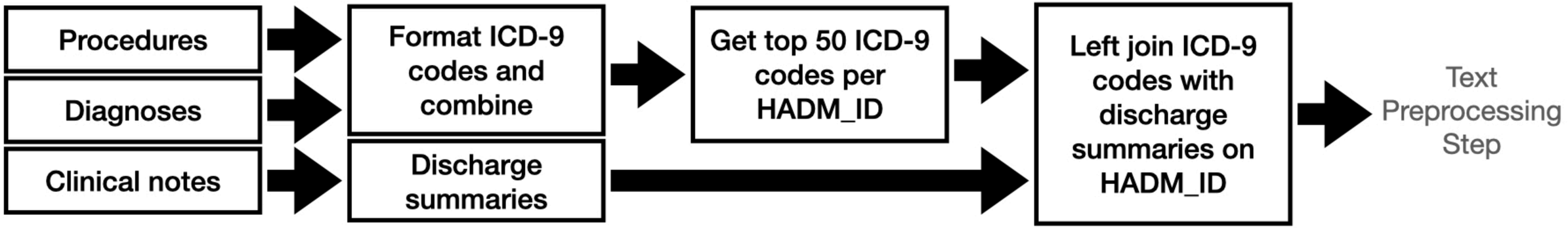
Clinical notes are text documents created by clinicians to document/record each patient encounter. It contains medical codes which effectively describes the diagnosis and treatment. Manually annotating these medical codes can be very labor intensive and prone to human error.

When medical codes are omitted, it becomes harder for others to understand the specific rationale behind certain specific diagnosis and treatment. As such, NLP (Natural Language Processing) techniques have been used to automate this labelling process.

The expected result is the reduction of time to label the documents leaving clinicians only for the validating process increasing reliability accuracy.

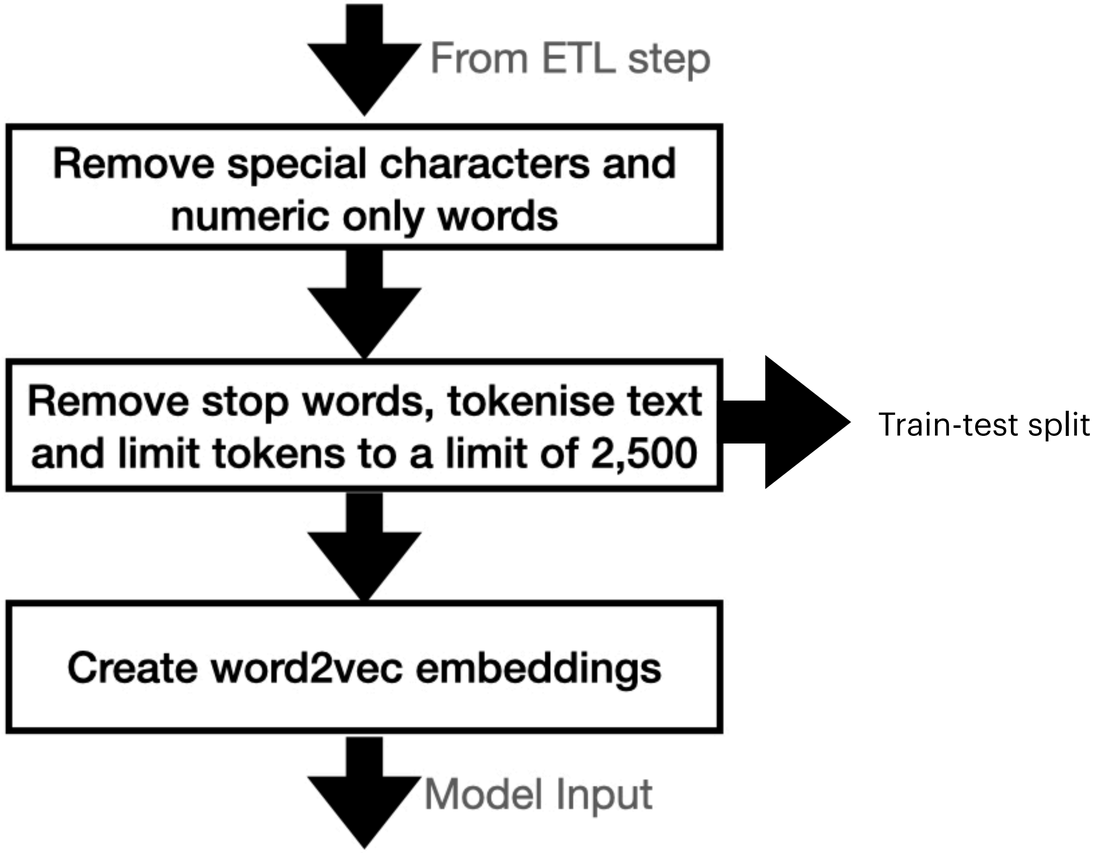
Data Pre-processing

ETL



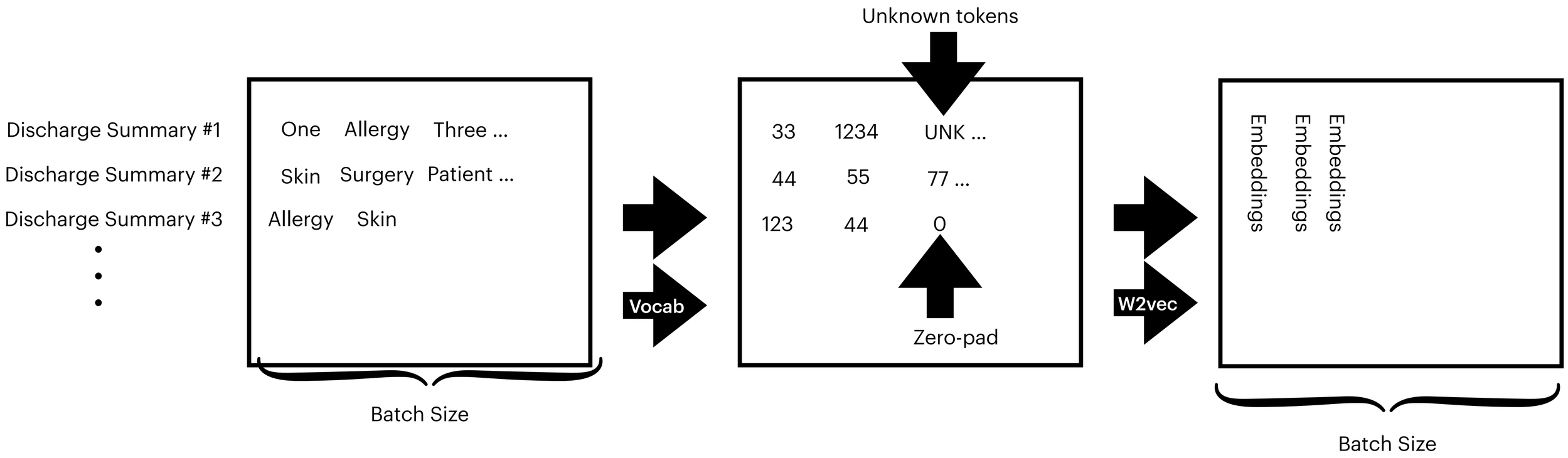
Data Pre-processing

Text-preprocessing



Data Pre-processing

Text-preprocessing(cont'd)



Proposed Model

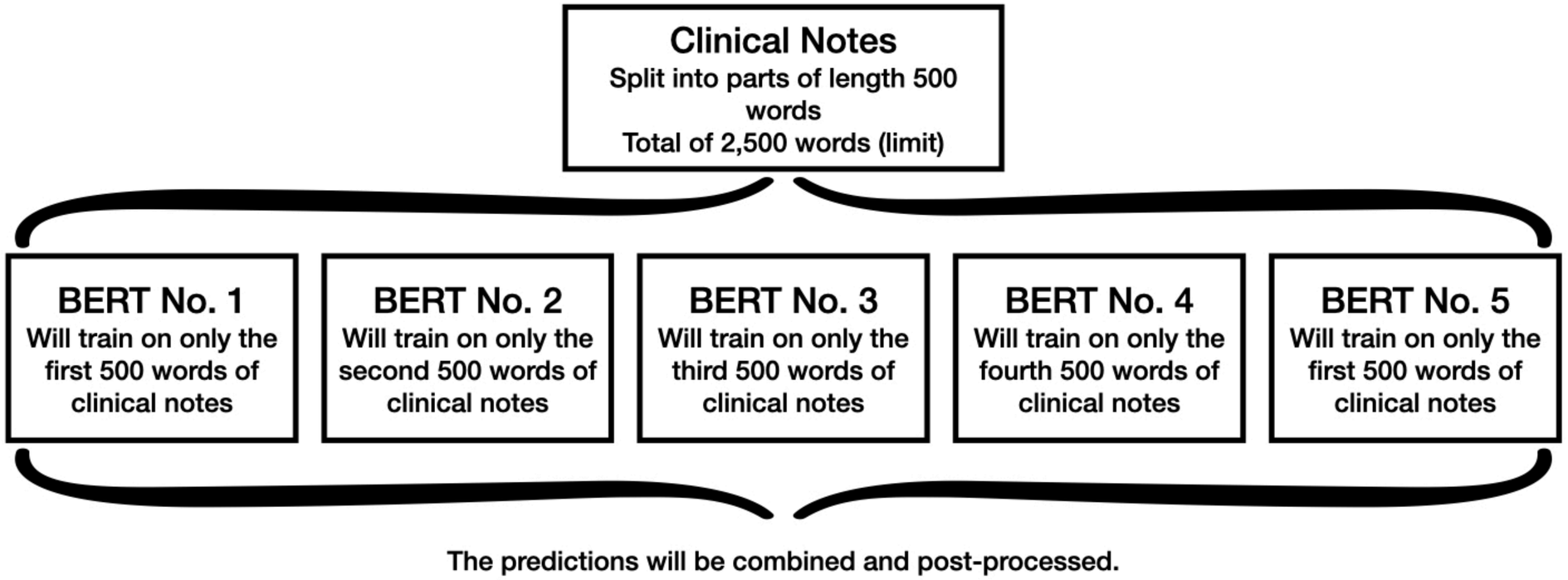
BERT

- BERT stands for Bidirectional Encoder Representations from Transformers
- Background
- BERT fine-tuned with MIMIC-III dataset to perform medical codes prediction

Data Pre-processing (BERT)

- The pre-processing pipeline is similar but instead of using the word2vec embedding, we make use of a BERT tokeniser.
- Essentially this tokeniser will make the input text compatible with BERT

Proposed Model



Baselines

- CBOW-LR
- CNN (Kim, 2014)
- BiGRU
- CAML (James Mullenbach, 2018)

Results

	AUC		F1		P@n	
Model	Macro	Micro	Macro	Micro	8	15
Logistic Regression	0.812	0.975	0.009	0.238	0.466	0.358
CNN	0.842	0.977	0.020	0.301	0.488	0.370
Bi-GRU	0.876	0.983	0.043	0.403	0.585	0.440
CAML	0.733	0.970	0.037	0.452*	0.592	0.455
BERT	0.801	0.846	0.285*	0.411	0.421	0.292

Conclusion

- Made use of a pre-trained language model BERT to perform ICD-9 code predictions from clinical notes.
- Evaluated the performance of BERT against LR, CNN, GRU and CAML

Future Work

- ICD-10 predictions
- BERT with LR/CNN/BiGRU/CAML
- Cater for non-standard writing and vocabulary
- Predict medical codes for future visits

Thank You!