

Predicting Patient Readmission Risk from Medical Text via Knowledge Graph Enhanced Multiview Graph Convolution

Qiu hao Lu¹, Thien Huu Nguyen¹, Dejing Dou¹²

¹University of Oregon

²Baidu Research

Readmission Prediction from Clinical Text

- Motivation
 - Readmission prediction models heavily rely on numerical and time-series features of patients
 - General text classification methods focus on locality and lack the ability to capture long-distance and non-consecutive semantics
 - External knowledge not leveraged
- Method
 - MedText: Graph-based text classification with knowledge graph as input

MedText

- Clinical text
 - Entity recognition and linking
 - **UMLS** subgraph
- Text-level graph representation
 - **Intra-Document**: co-occurrences of vertices within a sliding window
 - **Intra-UMLS**: shortest path lengths & string similarity
 - **Document-UMLS**: doc & entity embedding similarity

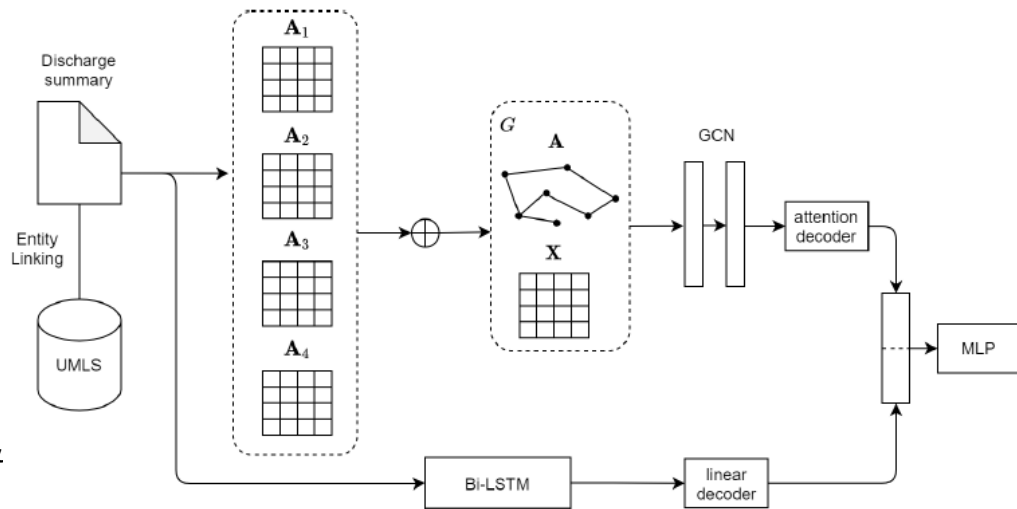


Figure 1: Architecture of MedText.

Experiments

- ICU readmission prediction
 - MIMIC-III Dataset
 - a large, freely-available database comprising deidentified health-related data associated with over 40,000 ICU patients, which includes information such as demographics, vital sign measurements made at the bedside (~1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports.
 - Metric
 - The Area under the Receiver Operating Characteristic (AUROC) as primary metric. It measures the overall performance of the recall with respect to different false positive rate.
 - AUROC is a commonly used metric for the readmission prediction task.
 - Along with some other metrics, including AUPRC, RP80.

Table 1: UMLS statistics.

Item	#
Entities (concepts)	2,983,840
Relations (general label)	14
Relations (additional label)	936
Triples	23,037,775

Experiments

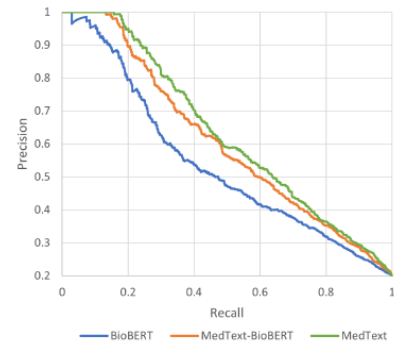
Table 2: Performance on patient readmission prediction.

Method	AUROC	AUPRC	RP80
BioBERT	0.775	0.538	0.200
MedText-BioBERT	0.811	0.610	0.278
ClinicalBERT	0.781	0.536	0.189
MedText-ClinicalBERT	0.812	0.615	0.277
CC-LSTM [21]	0.804	0.613	N/A
MedText	0.825	0.632	0.319

Table 3: Ablation analysis of MedText.

Method	AUROC	AUPRC	RP80
w/o $\mathcal{V}_1$	0.803	0.605	0.300
w/o $\mathcal{V}_{1,2}$	0.809	0.615	0.296
w/o $\mathcal{V}_{1,2,3}$	0.801	0.607	0.290
w/o $\mathcal{V}_{1,2,3,4}$	0.799	0.601	0.288
w/o D_T	0.808	0.601	0.275
Full	0.825	0.632	0.319

Figure 2: PR curve of MedText.



Potential Future Work

- Making medical predictions, e.g., readmission and mortality, based on pure textual data is promising. It is interesting to understand the different values between medical notes and numerical features in this prediction task.
- Another direction would be on the difference of clinical notes before and after hospitalization, which is interesting as most existing studies just use discharge summaries for prediction.