Towards a Health Knowledge Graph: Integration of External Knowledge into Pre-trained Language Models Capturing Physician Reasoning from EHR

Qiuhao Lu

Towards a Health Knowledge Graph

- There has been a surge of interest in applying machine learning techniques to clinical forecasting tasks.
    - mostly rely on numerical and time-series features of patients

- Extensive proliferation of EHRs
    - numbers are simple to handle but contain far less information than unstructured textual data, e.g., clinical notes in EHRs
    - motivates use to tackle the problem from NLP perspective

- Interpretability is a issue
    - one potential approach to improve interpretability is to infer a **health knowledge graph** from data

Integrating Knowledge into PLMs Capturing Physician Reasoning

- Multiple disparate sources of clinical knowledge exist
  - within EHR
    - documents, narratives, etc.
  - outside EHR
    - various external knowledge bases and ontologies, e.g., UMLS

- In order to fully construct such a health knowledge graph, ultimately these disparate knowledge repositories must be **consistently linkable** within a model.

- Motivates us to combine domain knowledge of disparate sources and formats under a unified framework, and inject them into PLMs.

Task

- Capturing physician reasoning for assigning diagnosis in the cardiovascular domain
    - by fine-tuning a language model on a diagnosis prediction task using physician comments as input

- To cope with domain knowledge of multiple sources and formats, we incorporate the **D**iverse **A**dapters for **K**nowledge **I**ntegration (DAKI) framework
    - the base models demonstrate decent performance on this predictive task.
    - when equipped with domain knowledge, performance gets consistently improved with an increased level of interpretability

Data Preparation

- Historical Clinical Notes
  - Event Types
    - Consult
    - Discharge Summary
    - Dismissal Summary
  - Section
    - Impression/Report/Plan

- Entity Linking
  - Map textual descriptions of diagnosis (free text, human written) to entities, e.g., UMLS CUIs

- Prepared <diagnosis, comment> pairs
  - Example

Patient	Diagnosis	Comment
11111111	Type 2 diabetes mellitus, diet controlled	Mr. XXXX's blood sugar this morning was 135. We will continue to monitor daily glucose and continue him on an 1800-calorie diabetic diet after his procedure later today.
11111111	Chronic kidney disease, stage 3	Mr. XXXX's creatinine today is 1.7. We will follow this closely throughout his hospital stay.
11111111	Organic heart disease	A. Coronary artery disease. B. Chronic atrial fibrillation. C. Dilated cardiomyopathy with decompensated heart failure. D. Status post CRT-D pacing with dislodged ICD and LV leads E. Status post tricuspid and pulmonary valve replacement with porcine valves The plan tomorrow is to proceed with a right ventricular ICD lead revision. This will likely require extraction of the current ICD lead with insertion of a new lead. Ideally, he would have revision of his coronary sinus lead as well, and this would require laser extraction of this lead as well. This procedure can be performed in the cardiac catheterization laboratory as per our protocol because his ICD lead has been in place for less than five years and his pacing lead for less than ten. The goals, risks, and alternatives of the procedure were discussed in great detail with the patient. The plan of care was discussed. A consent form was reviewed and signed. He has been typed and screened. We will keep him n.p.o. after midnight and be assured that his INR is not in a therapeutic range.

Label Distribution

- Statistics for CV texts (2015-01-01 ~ 2015-02-01, SciSpacy)
  - Number of documents: 428,209
  - Number of CV documents (service_code='CV---"): 16,450
  - Number of pairs: 18,177
  - Number of valid pairs: 17,099
    - 89 not recognized in label text
    - 989 recognized but not linked, i.e., 18177 = 89 + 989 +17099
- Statistics for CV texts (2015-01-01 ~ 2015-12-31, SciSpacy)
  - Number of documents: 4,931,788
  - Number of CV documents (service_code='CV---"): 192,360
  - Number of pairs: 185,896
  - Number of valid pairs: 174,980
    - 999 not recognized in label text
    - 9,917 recognized but not linked, i.e., 185896= 999 + 9917 +174980

Service Code	Service Description
CVD	Cardiovascular Diseases
CVDCL	Cardiovascular Diseases - Catherization Laboratory
CVDRHB	Cardiovascular Diseases - Cardiac Rehabilitation
CVDRYTM	Cardiovascular Diseases - Heart Rhythm
CVS	Cardiovascular Surgery

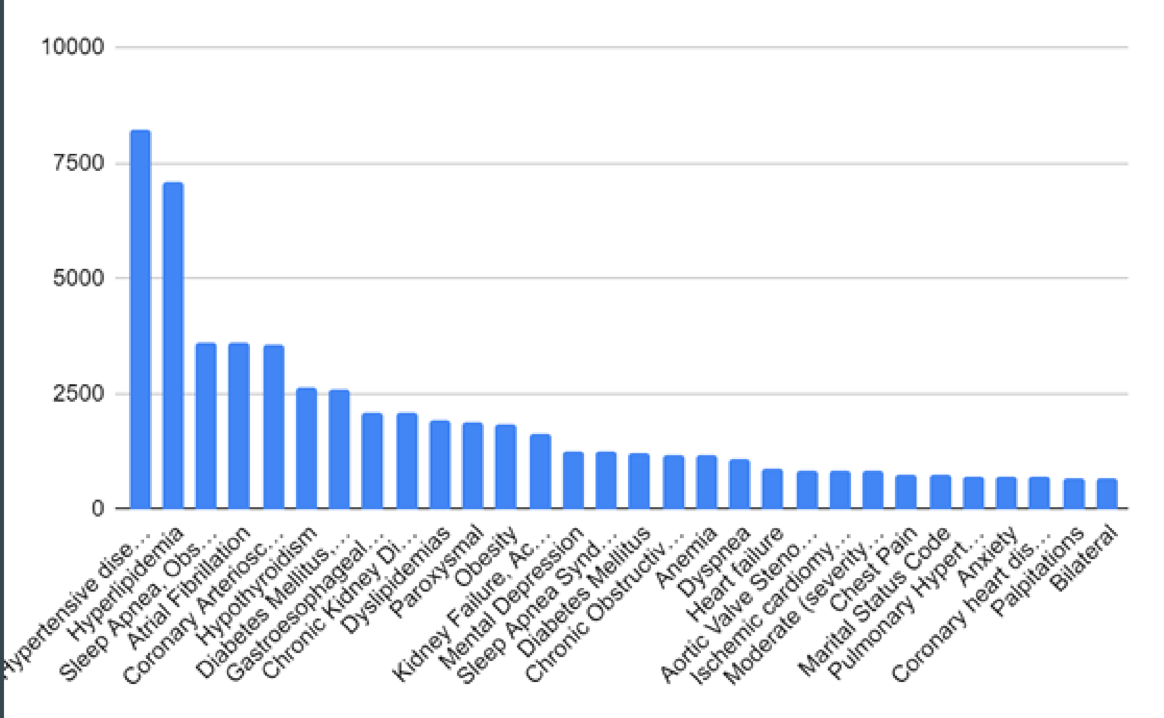
Label Distribution (UMLS)

- Label Sparsity
  - Most (linked) entities occur ONCE in the prepared dataset (1 month)
    - 93.89% entities frequency == 1 (16055/17099)
    - 98.45% entities frequency <= 10 (16837/17099)
  - 1 year
    - 97.76% entities frequency == 1
    - 99.21% entities frequency <= 10

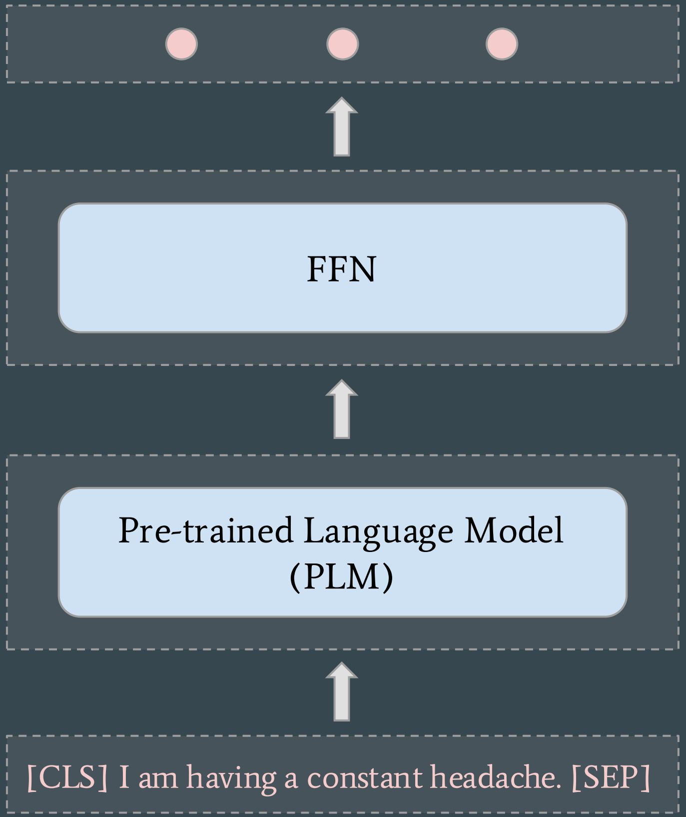
- Top-50 most frequent entities

CUI	counts	Name
C0020538	8227	Hypertensive disease
C0020473	7086	Hyperlipidemia
C0520679	3610	Sleep Apnea, Obstructive
C0004238	3585	Atrial Fibrillation
C0010054	3549	Coronary Arteriosclerosis
C0020676	2641	Hypothyroidism
C0011860	2619	Diabetes Mellitus, Non-Insulin-Dependent
C0017168	2094	Gastroesophageal reflux disease
C1561643	2080	Chronic Kidney Diseases
C0205178	1915	Dyslipidemias
C0242339	1876	Paroxysmal
C0205311	1829	Obesity
C0028754	1644	Kidney Failure, Acute
C0022660	1278	Mental Depression
C0011570	1244	Sleep Apnea Syndromes
C0037315	1225	Diabetes Mellitus
C0011849	1174	Chronic Obstructive Airway Disease
C0024117	1165	Anemia
C0002871	1075	Dyspnea
C0013404	900	Heart failure

Label Distribution: Top-30 Targets



Typical Classification Model



1) Pre-trained Language model
   a) BERT, BioBERT, ClinicalBERT, umlsBERT, SciBERT, ALBERT
   b) SciFive, ClinicalT5

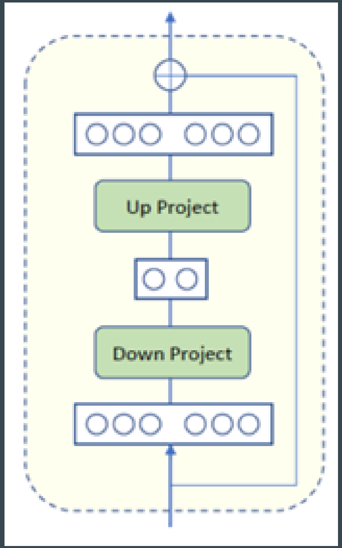
1) Inject Domain Knowledge into PLMs
   a) Diverse Adapters for Knowledge Integration (DAKI)

Adapters

- Adapters
  - simple & light-weight neural network
  - within a large pre-trained base model
    - For NLP, the base is usually a PLM, e.g., BERT
  - placement defines two paradigms

- Adapter training

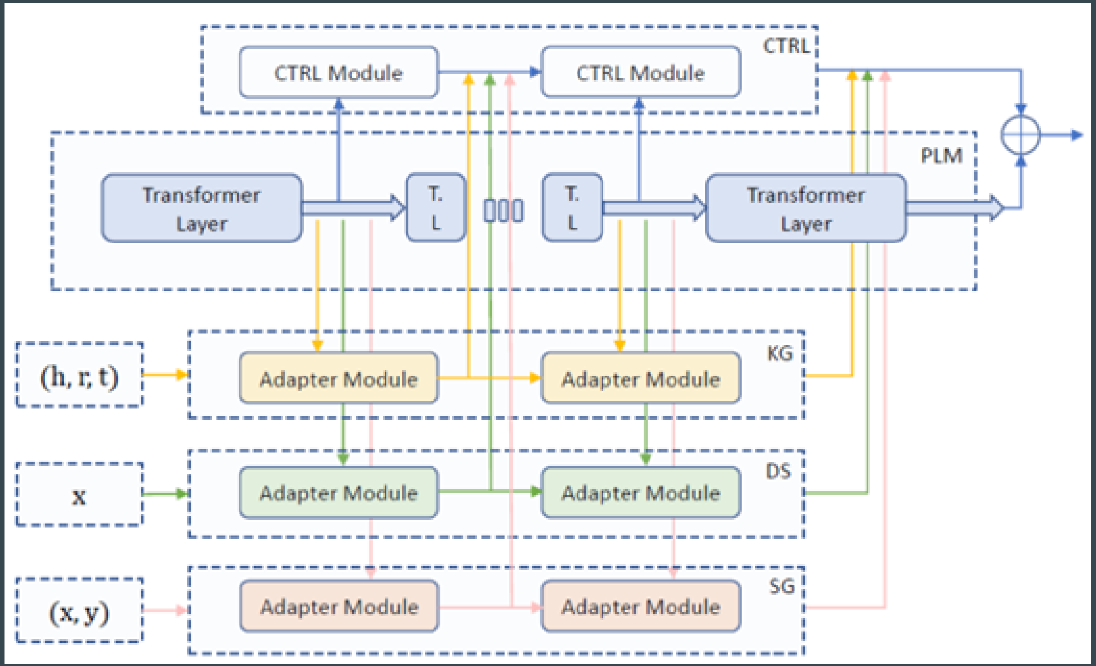
- Adapters instead of full model fine-tuning
  - Sequential fine-tuning and multi-task learning cause catastrophic forgetting and dataset imbalance
  - flexibility & parameter-efficiency



Adapter module.

Diverse Adapters for Knowledge Integration (DAKI)

- Base PLM
  - ALBERT-xxlarge-v2
- Pre-trained adapters
  - diverse sources
  - independently pre-trained
- Knowledge Controller
  - adaptively activate adapters
- Usage
  - Same as PLMs



Architecture of DAKI.

Adapters Pre-training - Knowledge Graph Adapter (KG)

- Link prediction
  - convert a triple (h, r, t) to a sequence and encode with DAKI

(diffuse adenocarcinoma of the stomach, disease has normal tissue origin, gastric mucosa)

↓

[CLS] diffuse adenocarcinoma of the stomach [SEP] disease has normal tissue origin [SEP] gastric mucosa [SEP]

$$\mathcal{L}_{\mathrm{KG}} = - \sum_{t \in \{\mathcal{T}^{+} \cup \mathcal{T}^{-}\}} \left(y \log \hat{y}_1 + (1 - y) \log \hat{y}_0 \right)$$

Adapters Pre-training - Disease Adapter (DS)

- Disease knowledge bridges the gap between disease terms and their textual descriptions.
  - e.g., (No history of blood clots or DVTs has never had chest pain prior to one week ago, Patient has **angina**) > > Entailment if PLM knows that angina refers to chest pain.

- Masked language modeling
  - randomly substitute 75% of the disease terms in the Wikipedia passages with [MASK] and optimize with a MLM objective

$$\mathcal{L}_{\mathrm{DS}} = \mathcal{L}_{\mathrm{mlm}}(T_{\Pi}|T_{\neg\Pi}) = -\frac{1}{K}\sum_{k=1}^{K}\log p(t_{\pi_k}|T_{\neg\Pi})$$

Adapters Pre-training - Semantic Grouping Adapter (SG)

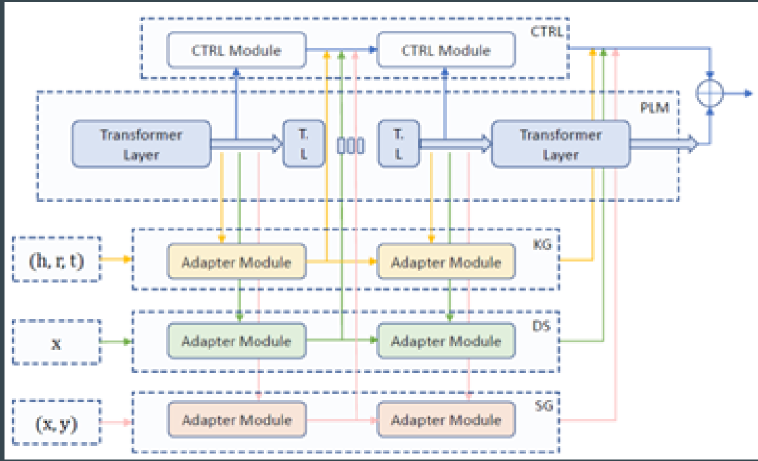
- Semantic grouping information helps pre-trained language models capture the connectivity between medical concepts, as well as between their descriptive texts
  - UMLS concept types are aggregated into 15 semantic groupings

- Multiclass classification

$$\mathcal{L}_{\text{SG}} = -\sum_{i=1}^{15} y_i \log \hat{y}_i$$

Knowledge Controller

- Adaptively integrate the knowledge adapters by assigning them different importance weights
  - As opposed to simple concatenation of the outputs of adapters, which is not intuitive, e.g., different tasks may require different knowledge.
- Knowledge controller is essentially a separate adapter with additional linear layers



$$\mathbf{Q}_i = \mathbf{W}_{Q_i}\mathbf{H}_{C_i} + \mathbf{b}_{Q_i}$$
$$\mathbf{K}_i = \mathbf{W}_{K_i}\mathbf{H}_{D_i} + \mathbf{b}_{K_i}$$
$$\mathbf{A}_i = \mathrm{softmax}(\mathbf{Q}_i\mathbf{K}_i^T)\mathbf{H}_{D_i}$$
$$\mathbf{Z}_i = \mathbf{W}_{V_i}\mathbf{A}_i + \mathbf{b}_{V_i}$$

Experiments

- Setup
  - input packed as sequence
  - DAKI, or PLM, as Encoder
  - Embeddings for [CLS] used for prediction

Adapter	Source	Size	Format
KG	UMLS Metathesaurus	$1,772,248$	(h, r, t)
DS	Wikipedia	$14,617$	x
SG	Semantic Network	$333,005$	(x, y)

Statistics of the datasets for pre-training KG, DS, SG. The formats are triples, passages, and textual definitions with labels, respectively

Experiments

- Data
  - Confirm target categories, e.g., remove History etc.
  - Patient level split
    - 80%/10%/10%

- Baselines
  - ALBERT-xxlarge-v2
  - BERT-base
  - ClinicalBERT-base

Experiments

Dataset	Test				Dev			
Metric(%)	Acc@1	Acc@3	Acc@5	Acc@10	Acc@1	Acc@3	Acc@5	Acc@10
BERT	48.81	69.70	76.73	85.37	50.56	70.21	76.84	84.92
ALBERT	48.02	69.35	76.70	85.82	50.47	69.37	76.00	84.60
ClinicalBERT	48.49	69.88	77.05	85.96	50.63	70.27	76.87	84.97
DAKI-BERT	48.20	70.26↑	77.64↑	86.38↑	50.59↑	70.19	76.77	84.78
DAKI-ALBERT	48.32↑	69.93↑	77.60↑	86.30↑	50.70↑	70.48↑	76.82↑	84.84↑
DAKI-ClinicalBERT	48.73↑	70.15↑	77.05	85.94	51.14↑	69.99	76.45	84.75

Acc@k is the number of times where the correct label is among the top k labels predicted (ranked by predicted scores).

Table 1: Overall performance.

Most & Least Successful Predictions

Target	F1
Marital Status Code	0.91
Sleeplessness	0.90
Gastroesophageal	0.82
Benign prostatic hypertrophy	0.82
Hypothyroidism	0.77
Mental Depression	0.71
Degenerative Polyarthritis	0.71
Constipation	0.69
Gout	0.67
Hypokalemia	0.66

(a) Most successful targets.

Target	F1
Mild Adverse Event	0.06
Coronary heart disease	0.07
Diabetes	0.07
Dyslipidemias	0.08
Chronic atrial fibrillation	0.08
Moderate	0.1
Cardiomyopathy, Dilated	0.19
Anticoagulation Therapy	0.2
Diabetes Mellitus	0.21
Paroxysmal	0.23

(b) Least successful targets.

DAKI-BERT on test set.

Biggest Impact on Least Successful Targets

Target	F1
Marital Status Code	0.91
Sleeplessness	0.90
Gastroesophageal	0.82
Benign prostatic hypertrophy	0.82
Hypothyroidism	0.77
Mental Depression	0.71
Degenerative Polyarthritis	0.71
Constipation	0.69
Gout	0.67
Hypokalemia	0.66

(a) Most successful targets.

Target	F1
Mild Adverse Event	0.06
Coronary heart disease	0.07
Diabetes	0.07
Dyslipidemias	0.08
Chronic atrial fibrillation	0.08
Moderate	0.1
Cardiomyopathy, Dilated	0.19
Anticoagulation Therapy	0.2
Diabetes Mellitus	0.21
Paroxysmal	0.23

(b) Least successful targets.

Target	ΔF1
Normocytic anemia	0.24
Anxiety	0.14
Bilateral	0.09
Non-ST EMI	0.09
Chronic atrial fibrillation	0.08
Coronary heart disease	0.07
Diabetes	0.07
Sleep Apnea Syndromes	0.06
Mild Adverse Event	0.06
Fatigue	0.05

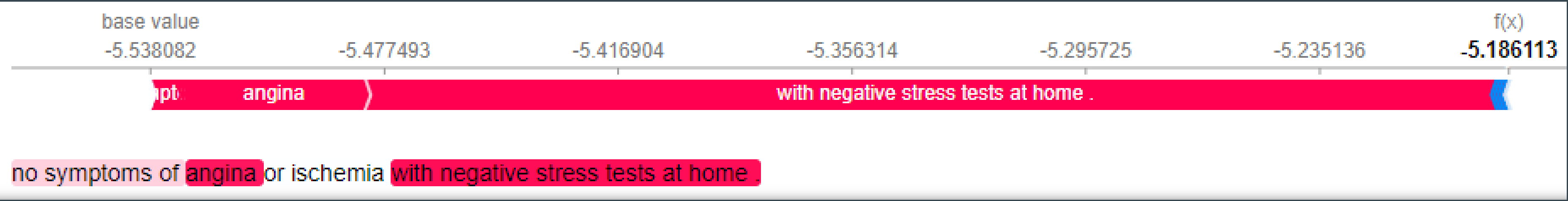
(c) Biggest impact.

Table 3: Impact pattern analysis. Red shows biggest impact on least successful targets.

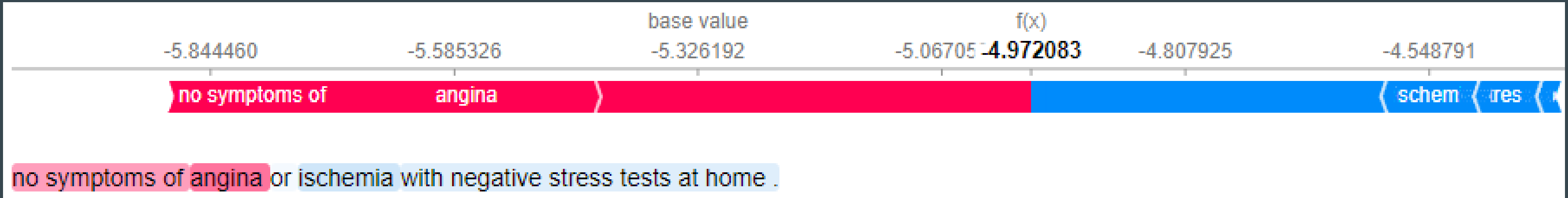
DAKI-BERT on test set.

Contribution Shift

ClinicalBERT

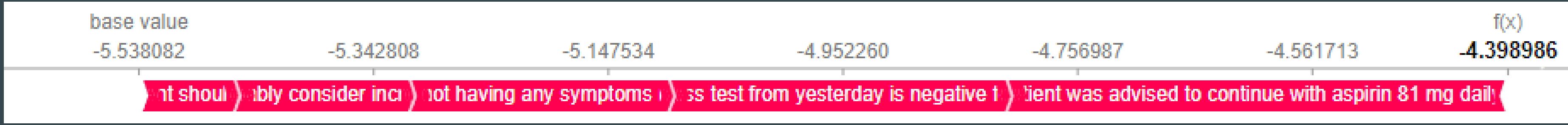


DAKI-ClinicalBERT



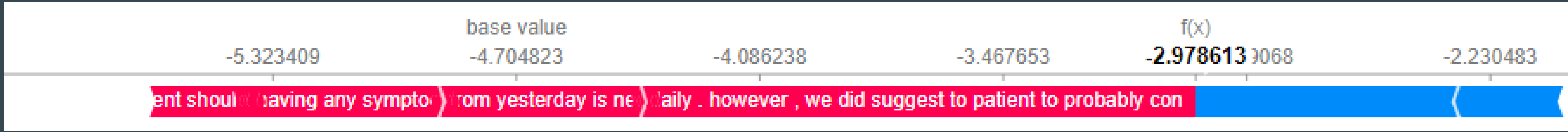
Contribution Shift

ClinicalBERT



patient is currently not having any symptoms related to his ca ##d . patient ' s stress test from yesterday is negative for is ##che ##mia . patient was advised to continue with as ##pi ##rin 81 mg daily . however , we did suggest to patient to probably consider increasing dose of lip ##itor back to 20 mg daily . from card ##iovascular health stand ##point , patient should be good to receive his f ##aa re ##cer ##ti ##fication .

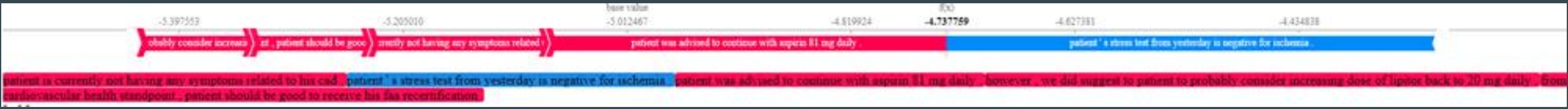
DAKI-ClinicalBERT



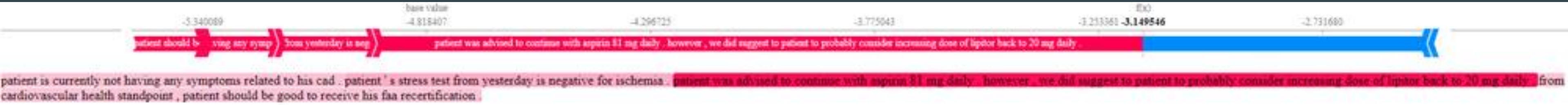
patient is currently not having any symptoms related to his ca ##d . patient ' s stress test from yesterday is negative for is ##che ##mia . patient was advised to continue with as ##pi ##rin 81 mg daily . however , we did suggest to patient to probably consider increasing dose of lip ##itor back to 20 mg daily . from card ##iovascular health stand ##point , patient should be good to receive his f ##aa re ##cer ##ti ##fication .

Contribution Shift

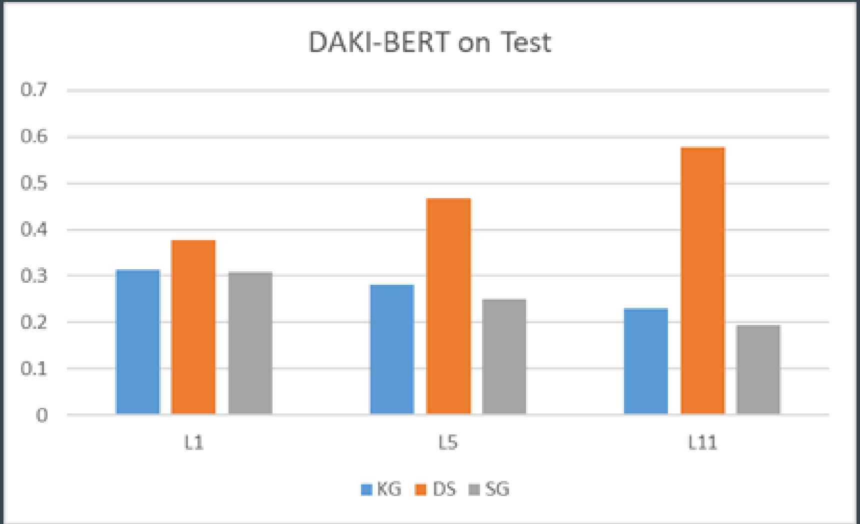
BERT

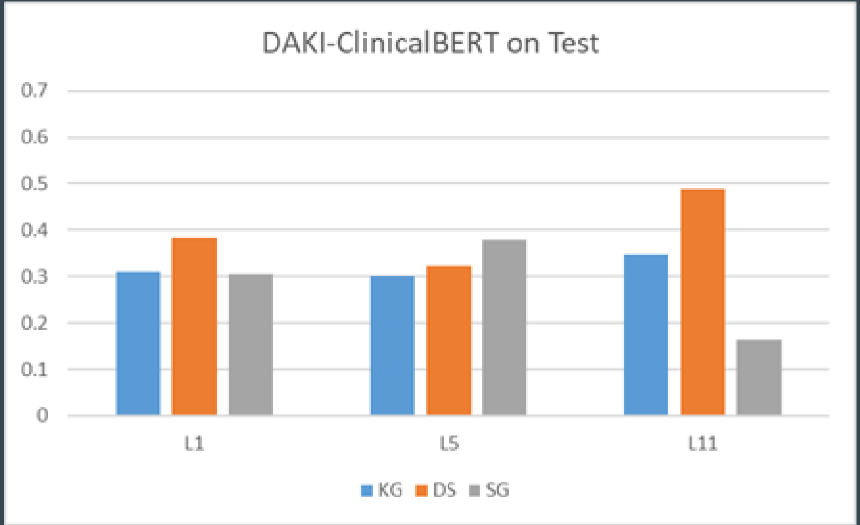


DAKI-BERT



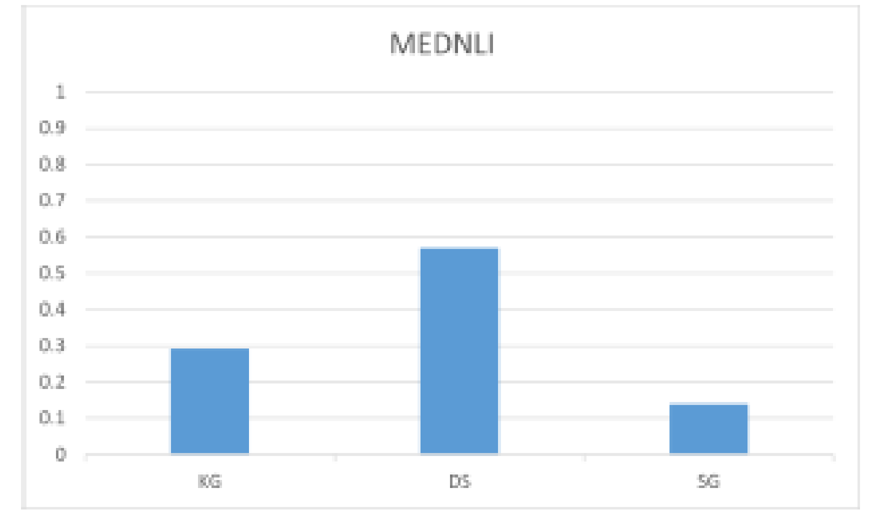
Knowledge Activation

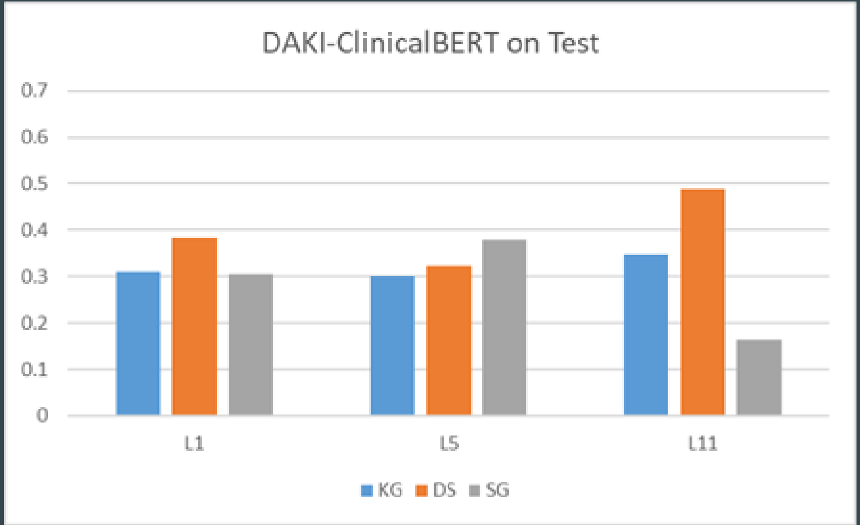




Activation levels of the adapters KG, DS, SG over the downstream tasks. We calculate the softmax activations in the 3 layers where the adapters are placed, and the activations are averaged over all instances in the test set.

Knowledge Activation - Similar Pattern





Activation levels of the adapters KG, DS, SG over the downstream tasks. We calculate the softmax activations in the 3 layers where the adapters are placed, and the activations are averaged over all instances in the test set.

Thank you!

What’s Next

- Knowledge adapter pre-training
  - CV-related
  - Ontology exploration

- Error analysis
  - Linking error (metamap)
  - Irrelevant categories (History, etc)
  - Discrepancy between comments and labels
  - accuracy for each diagnosis
  - Semantic type filtering
  - Top3 top5
  - SHapley Additive exPlanations (SHAP)

- writing