EMNLP 2024

DKEC: Domain Knowledge Enhanced Multi-Label Classification for Diagnosis Prediction

Xueren Ge, Abhishek Satpathy, Ronald Williams, John A. Stankovic, Homa Alemzadeh

Departments of Electrical and Computer Engineering and Computer Science

Contact: zar8jw@virginia.edu

University of Virginia Engineering · UVA DSA · UVA LINK LAB

Introduction

❖ Medic Notes based Diagnosis Prediction

- ➢ Emergency Medical Service (EMS)
- ➢ MIMIC-III ICD-9 Diagnosis Codes
- ➢ Long-tail Distribution
- ➢ Multi-Label Text Classification (MLTC)

❖ Existing Solutions

Models	Encoder	Attention Mechanism	Knowledge Integration	Knowledge Source	Datasets
(van Aken et al., 2021b)	BERT	Self-Attention	Pre-training	Wikipedia, PubMed	MIMIC-III
(Yang et al., 2022b)	MegatronBERT	Self-Attention	Pre-training	Wikipedia, PubMed	MIMIC-III
(Bolton et al., 2024)	GPT2	Self-Attention	Pre-training	PubMed	MedMCQA
(Mullenbach et al., 2018)	CNN	Label-wise Attention	ICD-9 hierarchy graph	ICD-9 description	MIMIC-III
(Rios and Kavuluru, 2018)	CNN	Label-wise Attention			MIMIC-III
(Li and Yu, 2020)	Multi-filter residual CNN	Label-wise Attention			MIMIC-III
(Zhou et al., 2021)	Multi-filter CNN	Shared Interactive Attention			MIMIC-III
DKEC (Ours)	Multi-filter CNN, Transformers	Label-wise Attention	Heterogeneous graph	Wikipedia, MayoClinic, ODEMSA	MIMIC-III & EMS

❖ Intuition: Domain Knowledge helps

- ➢ Compensate for data scarcity in fine-tuning
- ➢ Label relations can provide constraints in training

❖ Contribution

- ➢ Automated Knowledge Graph Construction by GPT-4
- ➢ Knowledge Graph incorporation with language models by heterogeneous label-wise attention
- ➢ DKEC outperform SOTAs on two real-world datasets

Knowledge Graph Construction

❖ Information Retrieval

- ➢ Description

❖ GPT-4 Chain-of-Thought (CoT) prompt

- ➢ Token Classification
- ➢ Span Detection
- ➢ Relation Extraction

❖ UMLS Concept Normalization

❖ Union of Knowledge Graphs from Multiple Sources

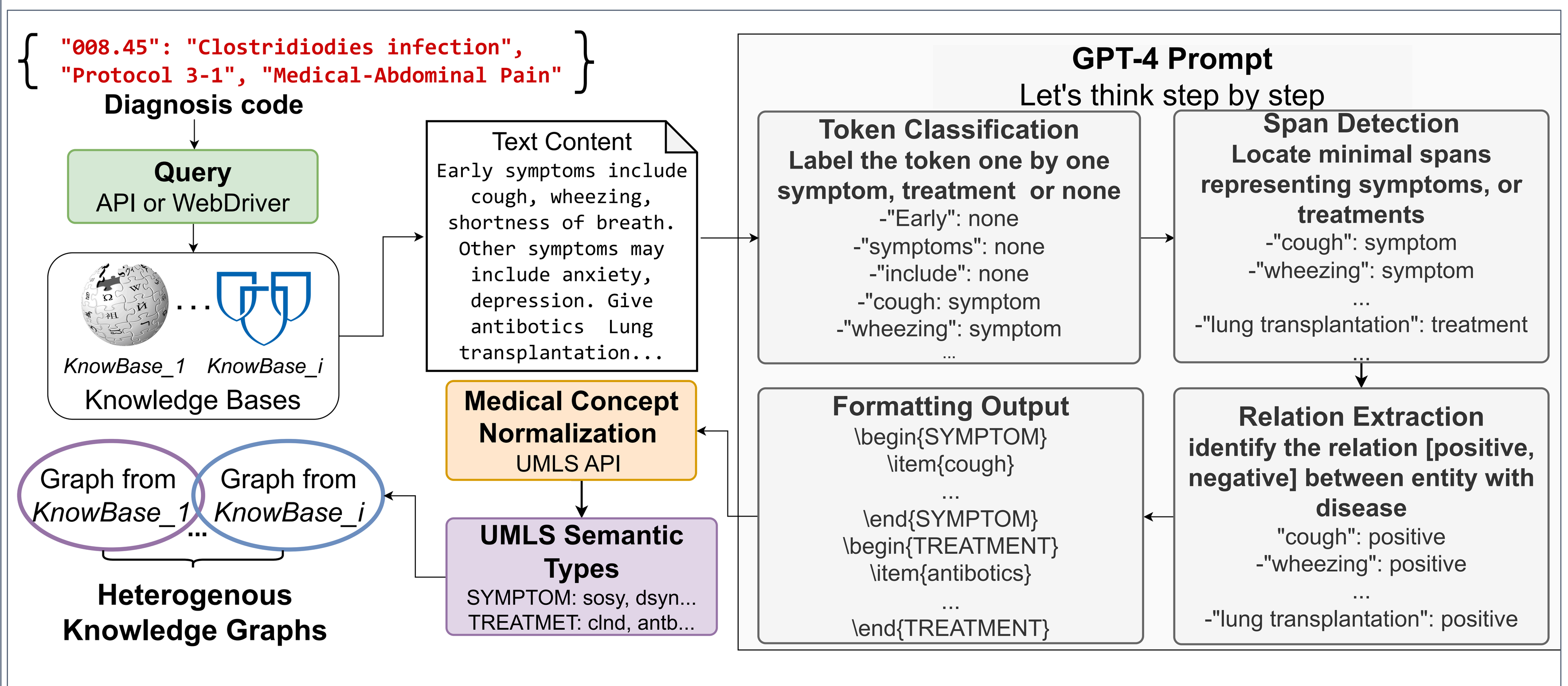


Figure 2: Knowledge Graph Construction

Knowledge Graph Quality Evaluation

❖ Symptoms and Treatments Manual Annotation

❖ 50 ICD-9 Diagnosis Codes

- ➢ Wikipedia and Mayo-Clinic

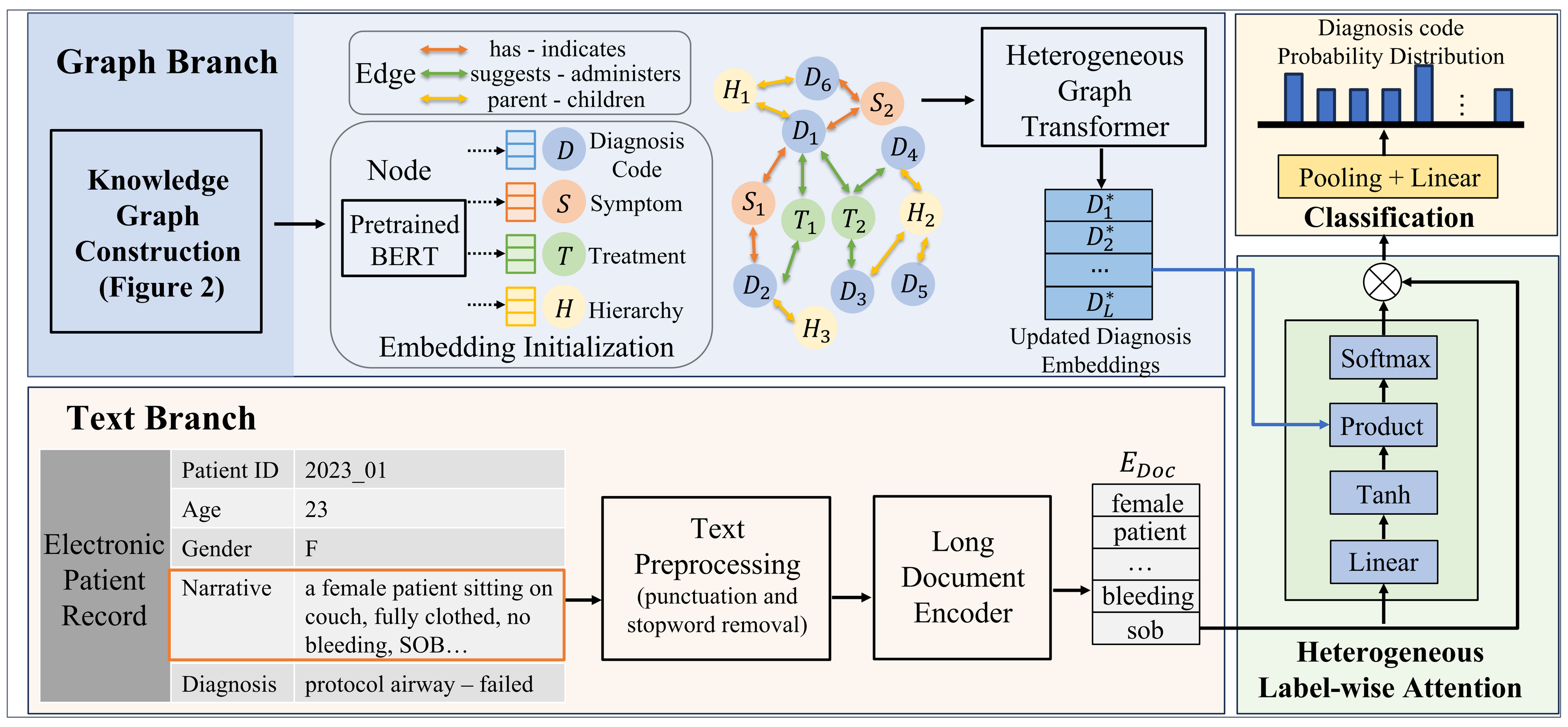
❖ 43 EMS Protocols

- ➢ ODEMSA

❖ One-shot CoT GPT-4 outperforms other baselines in medical entity extraction

wo/w NORM	Wikipedia (50 ICD-9 codes) Symptom	Wikipedia (50 ICD-9 codes) Treatment	Mayo Clinic (50 ICD-9 codes) Symptom	Mayo Clinic (50 ICD-9 codes) Treatment	ODEMSA (43 EMS protocols) Symptom	ODEMSA (43 EMS protocols) Treatment
MetaMap	47.62 / 51.53	34.66 / 41.95	44.83 / 49.12	41.82 / 46.44	41.34 / 43.61	39.20 / 41.95
cTAKES	48.74 / 52.58	36.01 / 43.35	42.60 / 46.67	39.67 / 45.35	38.02 / 42.47	48.96 / 52.31
ScispaCy	52.79 / 55.57	41.73 / 49.71	46.54 / 50.43	45.94 / 50.89	44.39 / 47.69	35.88 / 38.82
zero-shot GPT-4	51.99 / 58.77	17.93 / 32.13	52.98 / 63.37	26.16 / 36.48	76.07 / 79.72	10.17 / 23.50
one-shot CoT GPT-4	**84.63 / 86.57**	**85.70 / 89.12**	**82.03 / 86.72**	**90.43 / 93.90**	**86.96 / 91.01**	**86.48 / 88.92**

DKEC Pipeline



❖ Graph Branch

- ➢ Node: Diagnosis Code $D = \{D_k\}_{k=1}^{L}$; Symptom $S = \{S_k\}_{k=1}^{|S|}$; Treatment $T = \{T_k\}_{k=1}^{|T|}$; Hierarchy $H = \{H_k\}_{k=1}^{|H|}$
- ➢ Edge: Diagnosis Code - Symptom: $\overleftrightarrow{E_{DS}}$; Diagnosis Code - Treatment: $\overleftrightarrow{E_{DS}}$; Diagnosis Code - Hierarchy: $\overleftrightarrow{E_{DH}}$
- ➢ Heterogeneous Graph Transformer (HGT)

$$D^* = \mathrm{Linear}\big(\mathrm{HGT}(G)\big)$$

❖ Text Branch

$$E_{Doc} = \mathrm{Encoder}(Doc)$$

❖ Heterogeneous Label-wise Attention

- ➢ how informative medical document Doc is for all L labels

$$a_{Doc,k} = \mathrm{softmax}(\tanh(W_0 E_{Doc} + b_0) D_k^*)$$

$$A_{Doc} = [a_{Doc,1}\ a_{Doc,2}\ \dots\ a_{Doc,k}\ \dots\ a_{Doc,L}\ ]^T$$

$$E_{Doc}^{attn} = A_{Doc} E_{Doc}$$

❖ Classification

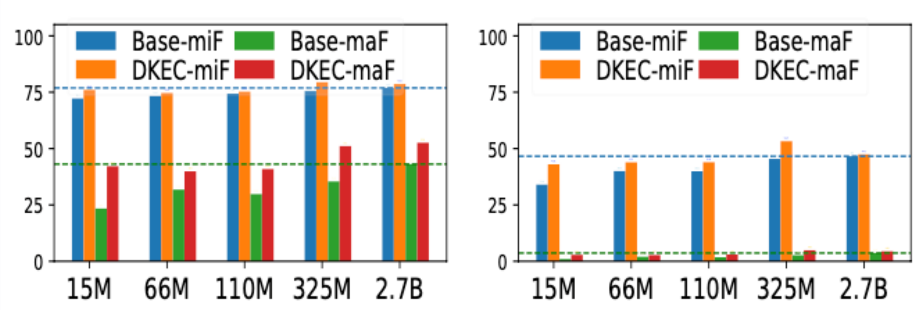
$$\hat{y}_{Doc} = \mathrm{Linear}(\mathrm{Pooling}(E_{Doc}^{attn}))$$

Results

❖ DKEC alleviates the class imbalance problem

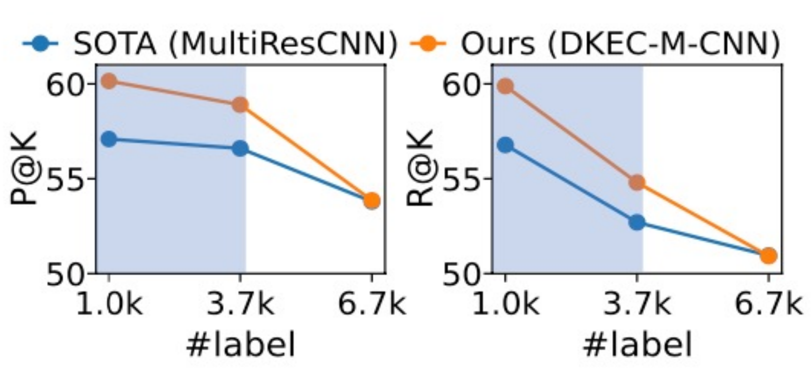
		Head Labels P@1	Head Labels R@1	Middle Labels P@1	Middle Labels R@1	Tail Labels P@1	Tail Labels R@1	Overall miF	Overall maF	Overall P@1	Overall R@1
EMS	CAML	78.6±1.3	77.7±1.3	33.0±0.5	32.6±0.6	22.7±4.5	22.7±4.5	63.7±1.2	22.4±1.3	65.0±1.6	63.5±1.5
	ZAGCNN	83.0±1.0	82.0±1.0	47.0±1.0	46.2±0.7	37.9±7.7	37.9±7.7	64.8±1.1	28.3±2.0	69.6±0.7	68.1±0.6
	MultiResCNN	84.3±0.2	83.2±0.2	35.6±1.8	35.0±2.0	25.0±2.3	25.0±2.3	65.8±0.2	26.1±0.5	67.9±0.3	66.3±0.3
	ISD	81.7±0.9	80.8±0.9	44.2±0.4	43.2±0.5	29.5±2.3	29.5±2.3	67.1±1.2	26.1±0.1	68.0±1.3	66.5±1.2
	GatorTron	89.4±0.5	88.4±0.5	66.0±0.4	64.7±0.7	57.1±2.2	57.1±2.2	75.5±0.6	35.4±1.9	77.3±0.6	75.4±0.6
	BioMedLM	89.3±0.3	88.2±0.3	71.3±0.7	70.1±0.6	47.6±4.3	47.6±4.3	76.9±0.7	43.1±1.7	78.4±0.6	76.6±0.6
	DKEC-M-CNN	85.2±0.7	83.0±0.7	53.2±1.3	52.7±1.1	45.1±2.1	45.1±2.1	68.6±0.4	32.4±0.6	72.4±0.4	71.7±0.6
	DKEC-GatorTron	**91.8**±0.1	**90.7**±0.1	**72.4**±0.4	**71.3**±0.4	**67.6**±2.3	**67.6**±2.3	**79.5**±0.5	**51.1**±1.5	**82.2**±0.5	**80.3**±0.6
		P@8	R@8	P@8	R@8	P@8	R@8	miF	maF	P@8	R@8
MIMIC-III	CAML	54.8±0.5	57.5±0.6	5.5±0.4	28.4±2.3	0.7±0.1	4.8±0.5	51.5±0.7	4.3±0.5	54.4±0.5	50.3±0.5
	ZAGCNN	55.3±0.2	58.0±0.2	6.6±0.1	34.4±0.7	1.8±0.1	11.7±0.8	52.1±0.4	4.0±0.3	55.2±0.2	51.2±0.3
	MultiResCNN	56.5±0.3	59.4±0.2	8.2±0.5	42.3±2.8	1.2±0.1	7.5±0.9	**55.6**±0.3	**6.0**±0.6	56.6±0.2	52.7±0.2
	ISD	51.8±0.5	53.8±0.5	6.1±0.2	31.7±1.2	1.9±0.2	12.6±0.9	46.8±1.3	2.8±0.2	51.6±0.5	47.5±0.5
	GatorTron	50.4±0.2	53.4±0.2	6.5±0.2	33.8±1.1	2.0±0.3	12.7±1.4	45.4±0.4	2.7±0.3	50.3±0.2	47.1±0.2
	BioMedLM	50.5±0.1	53.4±0.1	6.1±0.1	31.3±1.2	2.0±0.1	13.2±1.1	46.6±0.3	3.7±0.5	50.2±0.1	47.2±0.2
	DKEC-M-CNN	**58.6**±0.2	**61.5**±0.2	**9.6**±0.1	**49.2**±0.8	2.9±0.1	**19.2**±0.9	55.0±0.3	4.9±0.2	**58.9**±0.2	**54.8**±0.2
	DKEC-GatorTron	56.8±0.4	59.8±0.2	8.5±0.1	44.7±0.7	**3.1**±0.2	19.1±1.1	53.0±0.4	5.7±0.3	56.9±0.4	53.2±0.3

❖ DKEC enables smaller language models to achieve comparable performance to LLMs



(a) EMS (b) MIMIC-III

❖ DKEC maintains performance when external knowledge is available for all the labels



Acknowledgements

Richmond Ambulance Authority

NIST National Institute of Standards and Technology U.S. Department of Commerce

This work was supported by the award 70NANB21H029 from the U.S. Department of Commerce, National Institute of Standards and Technology (NIST)