Parameter-Efficient Domain Knowledge Integration from Multiple Sources for Biomedical Pre-trained Language Models

Qiuhao Lu[1], Dejing Dou[1,2], and Thien Huu Nguyen[1]

[1] University of Oregon
[2] Baidu Research

Biomedical Knowledge Integration for Pre-trained LMs

- (domain) pre-trained language models (PLMs) lack structured knowledge
  - PLMs are pre-trained on large unstructured free text, even domain variants
- Existing efforts for enriching PLMs with external knowledge rely on an auxiliary knowledge-driven training objective/regularization
  - costly full model fine-tuning
  - limited utilization of knowledge from multiple sources
  - focus on general domain knowledge while neglecting domain-specific knowledge

- To address these issues, we propose **DAKI.**
  - **D**iverse **A**dapters for **K**nowledge **I**ntegration

Adapters

- Adapter
  - 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 training
  - sequential fine-tuning and multi-task learning cause **catastrophic forgetting** and **dataset imbalance**
  - flexibility & parameter-efficiency

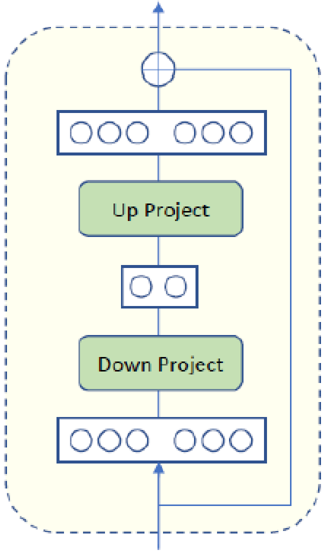

Figure 1: Adapter module.

DAKI

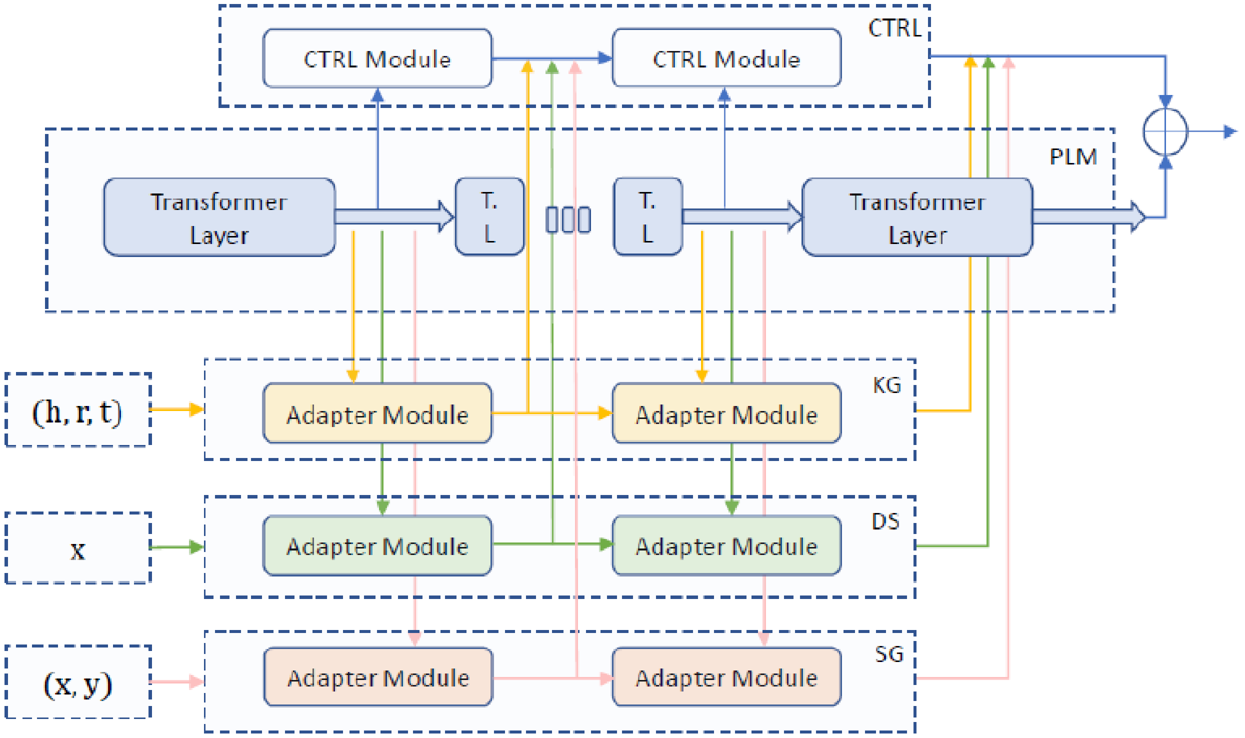

Figure 2: Architecture of DAKI.

Base PLM

- ALBERT-xxlarge-v2

Pre-trained adapters

- diverse sources
- independently pre-trained

Knowledge Controller

- attention-based
- adaptively activate adapters

Adapters Pre-training

- Knowledge-specific adapters
  - Knowledge Graph Adapter (KG)
  - Disease Adapter (DS)
  - Semantic Grouping Adapter (SG)

- Pre-training tasks
  - link prediction
  - masked language modeling
  - multiclass classification

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)

Table 1: Statistics of the datasets for pre-training KG, DS, SG. The formats are triples, passages, and textual definitions with labels, respectively.

Experiments

- Knowledge-driven biomedical NLP tasks
  - Question Answering (QA)
  - Natural Language Inference (NLI)
  - Named Entity Recognition (NER)

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

Experiments

Datasets	MEDIQA-2019			TRECQA-2017			MEDNLI	BC5CDR	NCBI
Metrics(%)	Accuracy	MRR	Precision	Accuracy	MRR	Precision	Accuracy	F1	F1
BERT	64.95	82.72	66.49	74.61	56.17	52.55	75.95	83.09	85.14
ClinicalBERT	67.30	84.78	70.59	77.00	52.56	56.62	81.50	84.90	87.25
SciBERT	68.47	84.47	68.07	77.23	54.57	57.54	80.94	86.16	87.24
BioBERT	68.29	83.61	72.78	77.12	49.84	57.25	81.86	85.99	87.70
diseaseBERT	66.40	83.33	68.94	75.33	56.41	54.01	77.29	83.47	86.81
umlsBERT	62.87	83.91	63.62	70.20	54.17	46.69	81.65	84.54	86.23
RoBERTa	72.49	86.74	74.67	75.33	51.76	54.01	81.65	83.04	85.83
ALBERT	76.54	88.46	81.41	75.09	58.57	53.03	85.48	84.28	87.56
diseaseALBERT	79.49	90.00	84.02	80.10	57.21	**62.40**	86.15	84.71	87.69
DAKI-BERT	69.47	85.06	70.17	77.95	54.65	58.27	77.85	83.43	85.67
DAKI-BioBERT	72.54	87.33	77.46	78.55	54.17	59.04	83.41	**86.51**	**89.01**
DAKI-RoBERTa	73.98	89.22	76.39	77.23	51.92	58.48	81.65	83.36	86.01
DAKI-ALBERT	**80.22**	**91.22**	**84.36**	**80.33**	**58.65**	62.31	**86.85**	84.86	87.86

Table 2: Performance of DAKI over downstream tasks.

Thank you!