Research Group Social Computing
TUM School of Computation, Information and Technology
Technical University of Munich

Knowledge-enhanced LMs in the biomedical domain

Daniel Homm, Mikhail Konov, and Johannes Burmester

Motivation

Documentation and data collection comprise a significant amount of physicians' time, which could be spent on patient care. "On average, 24% of working hours were spent on administrative duties". With high administrative workloads correlating with higher burnout rates and a tendency to see fewer patients [4].
We would like to lessen this burden with the help of language models that include knowledge about the biomedical domain.

Introduction

Using Adapters is a promising approach to infuse knowledge into a language model. They are small neural modules that are added between the Feed-forward layers of a transformer model. The weights of the base model are frozen, and only the Adapter weights are trained. This way, only 1-8% of the parameters of the network are updated. The approach we follow was originally introduced by [2] in Mixture-of-Partitions.

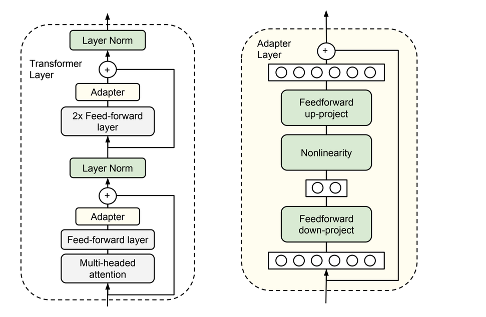


Figure 1 Adapter Module [5]

Methods

Knowledge Graph

- A knowledge graph is used to train the adapters (e.g. "Aspirin treats Headache"):

$$KG = \{(head, rel, tail) | head, tail \in Entities, rel \in Relations\}$$

- Unified Medical Language System (UMLS) [6]: Combines terminologies, relationships, and medical concepts into one knowledge graph
- Graph Partitioning with METIS [1] algorithm: coarsen, partition, refine

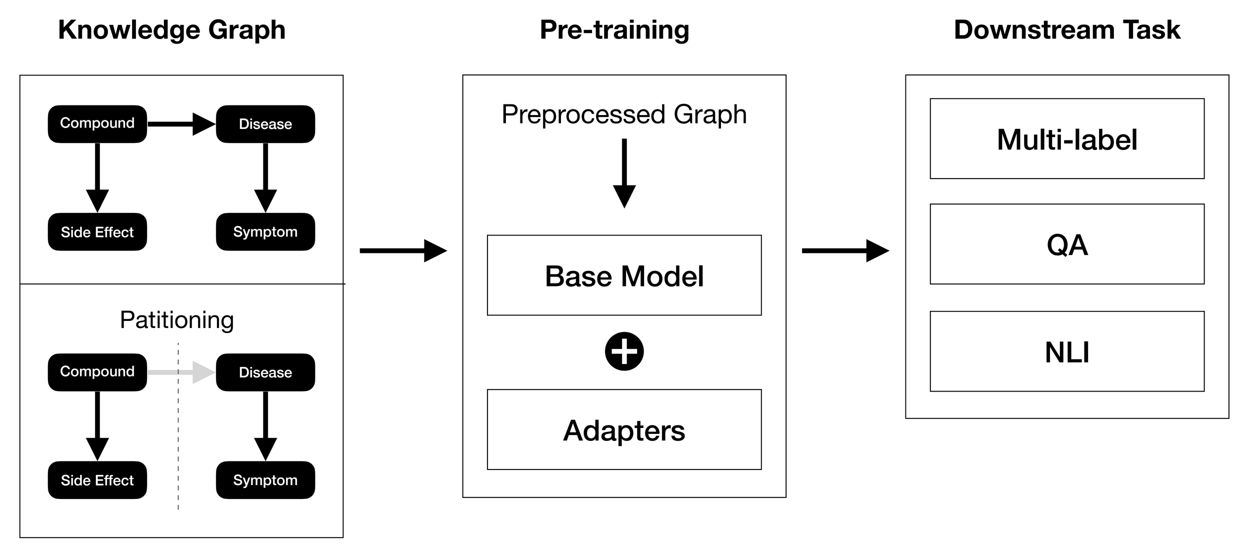


Figure 3 Overview of our Methods

Pre-training

- Edges of the graph are split into 20 fixed partitions.
- "Border edges": removed (NB) or included into both partitions.
- 2 pre-training approaches: link prediction (LP), entity prediction (EP). Both are modelled as multiclass classification task.
- For LP task it is feasible to pre-train on full dataset. 2 approaches: train single LP adapter (LP_FULL) or train multiple LP adapters that specialize in predicting certain relationship (LP_REL)

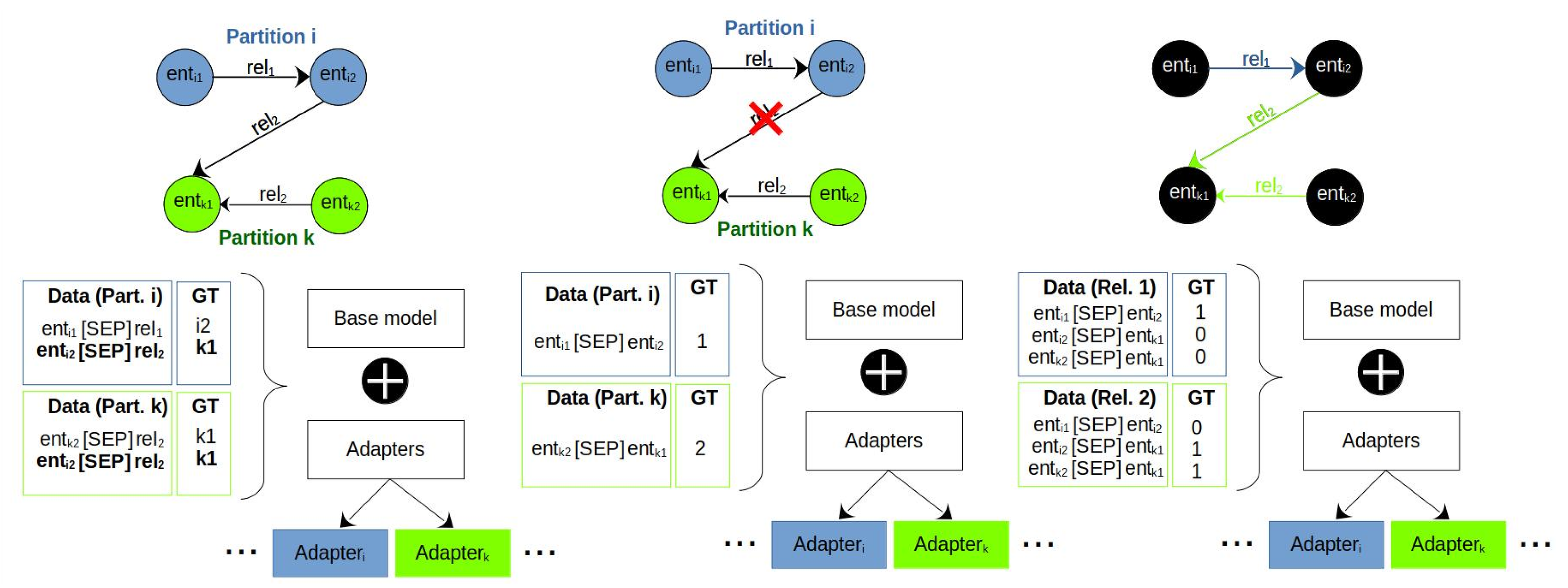


Figure 2 Schemas of pre-training approaches. Left-to-right: EP, LP_NB, LP_REL

Downstream Task

- PubMedQA as downstream task - 1000 expert-labelled QA pairs.
- Use adapter fusion to combine adapters. Fine-tune all weights of the final model.
- Use K-Cross Validation for robustness

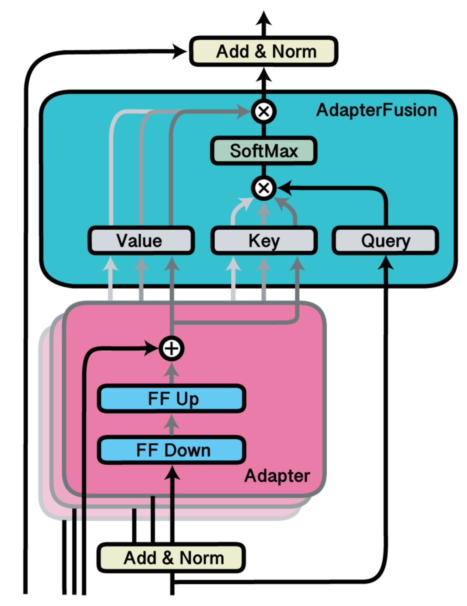


Figure 4 Adapter Fusion [3]

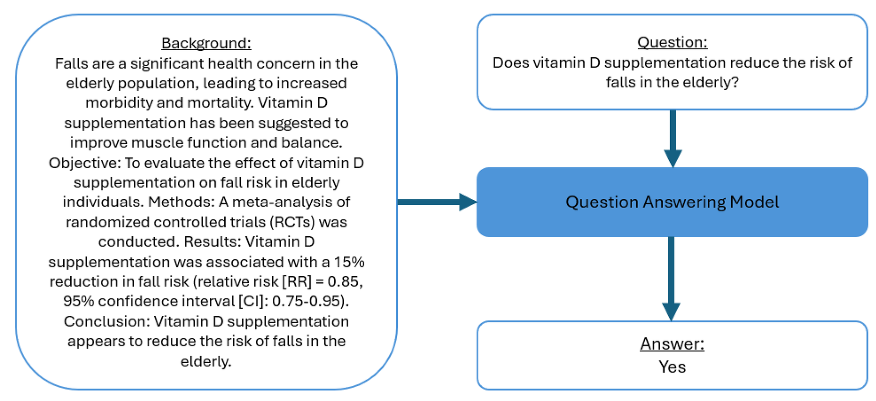
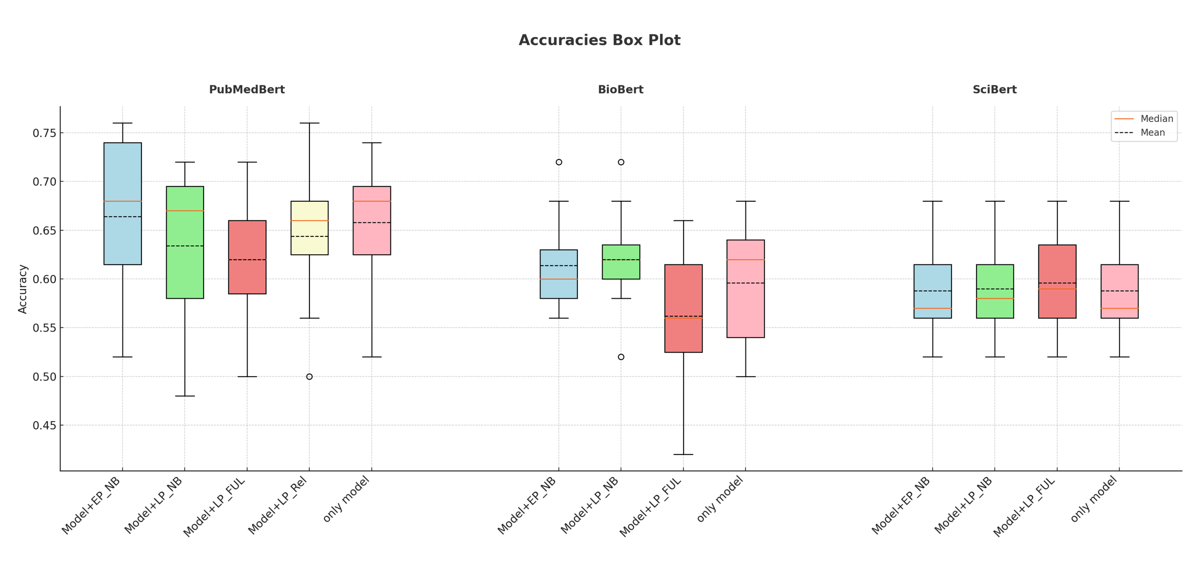
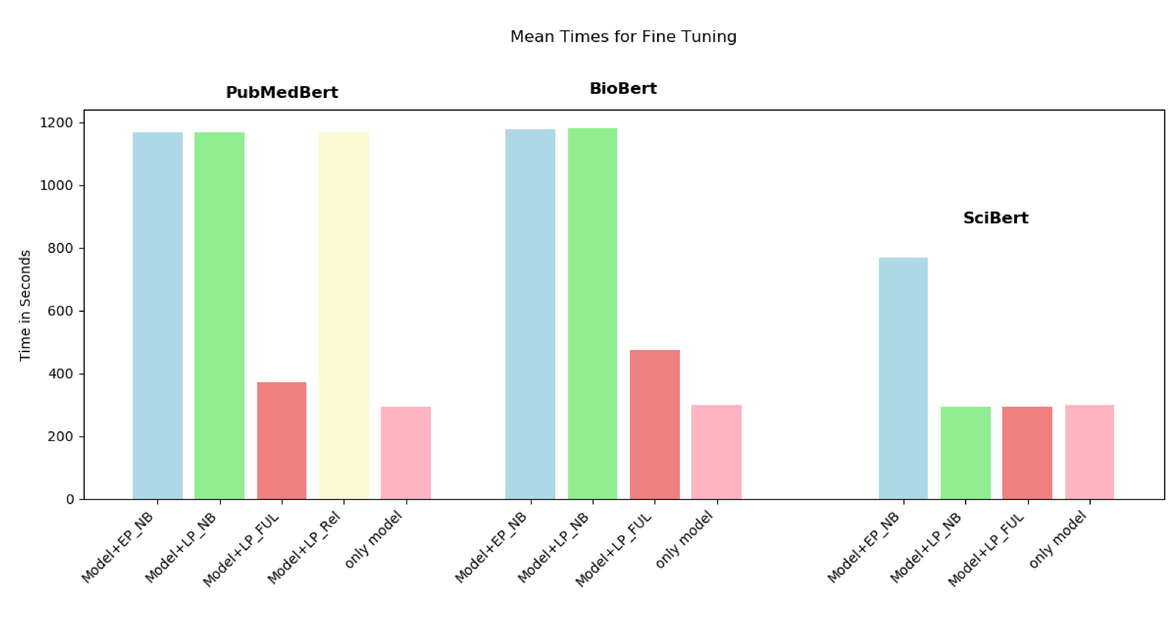


Figure 5 PubMedQA Example

Results





Discussion

- Fusion of multiple pre-trained adapters improves the quality of base model in most cases.
- Adapters have a limited learning capacity that negatively impacts performance on large datasets.
- Both LP and EP are trained on the same data, predicting different parts ⇒ pre-trained adapters result in similar behaviour

Conclusion

- Validation of the prior investigation of MOP [2] for model fine-tuning improvements with adapter fusion.
- Both LP and EP can outperform the baseline depending on the base model.
- Possible Future Topic: different adapter architectures, more advanced base models, more context from KG

[1] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on scientific Computing, 20(1):359–392,1998.
[2] Zaiqiao Meng, Fangyu Liu, Thomas Hikaru Clark, Ehsan Shareghi, and Nigel Collier. Mixture-of-partitions: Infusing large biomedical knowledge graphs into bert. arXiv preprint arXiv:2109.04810, 2021.
[3] Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. Adapterhub: A framework for adapting transformers, 2020.
[4] Sandhya K Rao, Alexa B Kimball, Sara R Lehrhoff, Michael K Hidrue, Deborah G Colton, Timothy G Ferris, and David F Torchiana. The impact of administrative burden on academic physicians: results of a hospital-wide physician survey. Academic Medicine, 92(2):237–243, 2017.[1] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on scientific Computing, 20(1):359–392, 1998.
[5] Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp, 2019
[6] Olivier Bodenreider. The unified medical language system (umls): integrating biomedical terminology. Nucleic acids research, 32(suppl$_1$) : $D267^{\sim}D270, 2004.$