Coarse-to-fine Knowledge Graph Domain Adaptation based on Distantly-supervised Iterative Training

1st Wenxiong Liao
South China University of Technology
GuangZhou, China
cswxliao@mail.scut.edu.cn

2nd Zhengliang Liu
University of Georgia
Athens, USA
zl18864@uga.edu

3rd Yiyang Zhang
South China University of Technology
GuangZhou, China
zyyinyourarea@163.com

4th Xiaoke Huang
South China University of Technology
GuangZhou, China
csxkhuang@mail.scut.edu.cn

5th Fei Qi
South China University of Technology
GuangZhou, China
fqiscut@foxmail.com

6th Siqi Ding
South China University of Technology
GuangZhou, China
d15995291636@163.com

7th Hui Ren
Massachusetts General Hospital and Harvard Medical School
Boston, USA
hren2@mgh.harvard.edu

8th Zihao Wu
University of Georgia
Athens, USA
zihao.wu1@uga.edu

9th Haixing Dai
University of Georgia
Athens, USA
hd54134@uga.edu

10th Sheng Li
University of Technology Virginia
Blacksburg, USA
vga8uf@virginia.edu

11th Lingfei Wu
University of Georgia
Athens, USA
Teddy.lfwu@gmail.com

12th Ninghao Liu
University of Georgia
Athens, USA
ninghao.liu@uga.edu

13th Quanzheng Li
Massachusetts General Hospital and Harvard Medical School
Boston, USA
li.quanzheng@mgh.harvard.edu

14th Tianming Liu
University of Georgia
Athens, USA
tliu@uga.edu

15th Xiang Li *
Massachusetts General Hospital and Harvard Medical School
Boston, USA
xli60@mgh.harvard.edu

16th Hongmin Cai
South China University of Technology
GuangZhou, China
hmcai@scut.edu.cn

Abstract—The knowledge graph (KG) is a highly needed basis to support the high-fidelity and high-interpretablity modeling of various tasks in healthcare artificial intelligence. In this work, we focus on constructing an oncology knowledge graph that will be used in downstream cancer research and solution development. Modern supervised learning for knowledge graph construction requires a large amount of manually labeled data, which makes the process time-consuming and labor-intensive. Although there exists multiple research on named entity recognition and relation extraction based on distantly supervised learning, constructing a domain-specific knowledge graph from large collections of textual data without manual annotations is still an urgent problem to be solved. In response, we propose an integrated framework for adapting and re-learning knowledge graphs from a general domain (biomedical in our case) to a fine-defined domain (oncology). In this framework, we apply distant-supervision on cross-domain knowledge graph adaptation. Consequently, no manual data annotation is required to train the model. We introduce a novel iterative training strategy to facilitate the discovery of domain-specific named entities and triplets. Experimental results indicate that the proposed framework can perform domain adaptation and construction of knowledge graphs efficiently.

Index Terms—Knowledge Graph Domain Adaptation, Knowledge Graph Construction, Named Entity Recognition, Relation Extraction

I. INTRODUCTION

In healthcare, the development of robust and interpretable clinical decision support systems and the corresponding research requires both a substantial amount of data and effective modeling of the medical domain knowledge [1]. Knowledge
graphs (KG) have been explored in healthcare to represent the underlying relationships of domain knowledge [2], [3].

However, existing approaches are not effective for constructing domain-specific KGs, especially for oncology, where limited access to oncological expertise hinders the supply of labeled training data. Insufficient labeled data typically leads to suboptimal performance. In fact, the dependence on sizeable training data significantly diminishes the real-world potential of data-driven KG construction methods based on supervised learning. In addition, although rule-based methods (based on resources such as Stanford CORE NLP) do not have stringent data demands, they typically suffer from suboptimal hand-crafted feature designs and the absence of helpful fine-grained connections to the domain data. Consequently, automatically constructing knowledge graphs directly from natural texts has attracted close attention in recent years [4], [5].

In order to address these challenges, we investigate the coarse-to-fine learning for constructing an oncology knowledge graph that leverages knowledge from general biomedical KGs, especially the distantly-supervised interactive training to achieve knowledge graph domain adaptation. In distantly-supervised learning, fine-domain KGs are derived from the general-domain KGs. For example, a biomedical KG that covers broad concepts and common sense knowledge in the biomedical domain can serve as the base KG for a specialized oncology KG. Therefore, the KG in the coarse domain can be used as a knowledge base for distant supervision, thus avoiding the need for extensive manual annotations. However, only using the KG of the coarse domain as the knowledge base might limit the model’s ability to discover domain-specific named entities and triplets in the fine domain, further limiting the construction of the fine domain KG. Thus in this paper, we propose a novel coarse-to-fine knowledge graph domain adaptation (KGDA) framework. Our KGDA framework utilizes an iterative training strategy to enhance the model’s ability to discover fine-domain entities and triplets, thereby facilitating fast and effective coarse-to-fine KG domain adaptation.

Overall, the contributions of our work are as follows:

- An integrated framework for adapting and re-learning KG from coarse-domain to fine-domain is proposed. Our framework does not require human annotated samples with distant-supervision for cross-domain KG adaptation, and the iterative training strategy is applied to discovering domain-specific named entities and new triplets.
- The proposed method can be adapted to various pre-trained language models (PLMs) and can be easily applied to different coarse-to-fine KGDA tasks. It is so far the simplest data-driven approach for learning a KG from free text data, with the help of the coarse domain KG.
- Experimental results demonstrate the effectiveness of the proposed KGDA framework. The source code of this paper has published in https://github.com/WenxiongLiao/KGDA.

II. METHODOLOGY

A. Notation and task definition

An unstructured sentence \( s = [w_1, w_2, w_3, ..., w_n] \) indicates a sequence of tokens, where \( n \) is its length. A dataset \( \mathcal{D} \) is a collection of unstructured sentences (i.e. \( \mathcal{D} = \{s_1, s_2, s_3, ..., s_m\} \)). The knowledge graph, denoted as \( \mathcal{K} \), is a collection of triplets \( t = (e_i, r_j, e_k) \), where \( e_i \in \mathcal{E} \) and \( e_k \in \mathcal{E} \) are the head entity and the tail entity respectively, and \( r_j \in \mathcal{V} \) is the relation between \( e_i \) and \( e_k \). Here we denote coarse-domain KG as \( \mathcal{K}_c \) and fine-domain KG as \( \mathcal{K}_f \).

In a typical scenario of KG domain adaptation, we have an existing coarse-domain KG and a large amount of unlabeled text in the fine domain. For example, when constructing an oncology KG, we can utilize the existing biomedical KG and collect oncology-related literature as unlabeled text. KG constructed from the fine domain data would include overlapping triplets with the coarse-domain KG and new triplets representing domain-specific knowledge. Specifically, the fine-domain KG contains the following three types of triplets:

- **Overlapping triplets** \( \mathcal{T}_O \): Triplets that also existed in the coarse-domain KG, indicating knowledge overlapping between the coarse and fine domains.
- **Triplets of new relations but overlapping entities** \( \mathcal{T}_R \): Triplets with both entity pairs existing in the coarse-domain KG but no indicated relationship.
- **Triplets of new entities** \( \mathcal{T}_E \): Triplets with at least one entity not existing in the coarse-domain KG. Consequently, the relationship is also unknown in the coarse domain.

Both \( \mathcal{T}_R \) and \( \mathcal{T}_E \) belong to the specific knowledge of the fine domain. The goal of the coarse-to-fine KGDA task is to adapt the KG from the coarse domain to the fine domain and leverage the knowledge from the coarse domain to guide the mining of new knowledge specific to the fine domain.

B. Iterative training framework

While it is trivial to identify the overlapping entities \( \mathcal{E}_O \) and triplets \( \mathcal{T}_O \) by distant supervision, if the NER and RE models are trained on the entire corpus, they will not be able to recognize the fine domain-specific named entities and triplets (\( \mathcal{T}_R \) and \( \mathcal{T}_E \)). Because the distant-supervision labels are generated by matching \( \mathcal{K}_c \). Thus we introduce an iterative training strategy to construct \( \mathcal{T}_R \) and \( \mathcal{T}_E \) from the text and adapt the knowledge from \( \mathcal{K}_c \) to \( \mathcal{K}_f \).

The overall framework of the iterative training scheme is shown in Fig. 1, and Algorithm 1. Rather than performing distant-supervision training on the whole unlabeled text corpus, the core mechanism of the proposed iterative training is to split the whole unlabeled dataset into \( n \) sub-datasets without intersection. Before building distant-supervision corpus, the trained model is used to predict the text corpus for getting specific knowledge of fine-domain, which is conducive to mining \( \mathcal{T}_R \) and \( \mathcal{T}_E \) of the fine-domain.

As shown in Figure 1, firstly, it is necessary to preprocess the acquired text corpus in the fine domain. Preprocessing
operations include: handling special characters, word segmentation, filtering sentences using human-defined rules (such as sentence length), etc. Then, our framework involves two neural network models: NER model and RE model. We replace the PLM’s output layer with a classifier head as NER model $model_N$ and fine-tune it by minimizing the cross-entropy loss on distant-supervision NER corpus. Additionally, we apply the BIO scheme [6] to generate NER sequence labels. For the RE task, we use the template to generate distant-supervision samples. The template we adopted is "[CLS] head entity (head entity type) [SEP] tail entity (tail entity type) [SEP] sentence". The RE model $model_R$ is defined as a PLM with a relation classifier. The feature of special token [CLS] fed into sentence $(constructed by either schemes) to the total sample size. The parameter $ratio_n$ controls the ratio of negative samples (constructed by either schemes) to the total sample size. The parameter $ratio_o$ controls the ratio of entity pairs constructed by the second scheme (i.e., via sampling the words outside the domain) to the size of negative samples, respectively.

In addition to the $K_c$ in the source domain, we use $K_c$, $E_{conf}$, and $T_{conf}$ as knowledge bases for constructing the remotely supervised corpus. This would ensure that the NER and RE models can identify the overlapping knowledge between $K_c$ and $K_f$, while at the same time be guided to discover the new knowledge specific to the fine domain.

As shown in Algorithm 2, for building the distantly-supervised NER corpus $corp_N$, the sentence $D_i^d$ is firstly string-matched with the knowledge bases $K_c$ and $E_{conf}$ to get the entities $entities_k$ based on $K_c$ and the entities $entities_c$ based on $E_{conf}$ in the sentence (line 5). Afterward, the matched entities are merged into overlapping entities $E_O$, and the NER label sequences are generated through the BIO strategy to merge into $corp_N$ (line 7). For building the distantly-supervised RE corpus $corp_R$, we firstly take $K_c$ and $T_{conf}$ as knowledge bases and use entity pair matching to match the triplets $triplets_c$ based on $K_c$ and the triplets $triplets_c$ based on $T_{conf}$ appearing in the sentence $D_i^d$ (line 8). We then
Algorithm 1 Iterative training KGDA framework

Input: Text corpus \(D = \{D_1, D_2, ..., D_n\}\), coarse-domain KG \(K_c\), out-of-domain words \(W_O\)

Parameter: Initialized NER model \(model_N\), initialized RE model \(model_R\)

Output: fine-domain kg \(K_f\)

1: Let new entities \(E_{new} = \{\}\) , new entities with high confidence \(E_{conf} = \{\}\) , new triplets \(T_{new} = \{\}\) , new triplets with high confidence \(T_{conf} = \{\}\) .
2: \(corp_N, corp_R, E_O, T_O = \text{get_distant_corpus}(D_1, K_c, E_{conf}, T_{conf})\)
3: train_NER(model_N, corp_N)
4: train_RE(model_R, corp_R)
5: \(i = 2\)
6: while \(i <= n\) do
7: \(E_{new}, E_{conf}, T_{new}, T_{conf} = \text{get_specific_knowledge}(D_i, K_c, E_{new}, E_{conf}, T_{new}, T_{conf})\)
8: \(corp_N, corp_R, E_O, T_O = \text{get_distant_corpus}(D_i, K_c, E_{conf}, T_{conf}, W_O)\)
9: \(corp_N = corp_N \cup corp_{c}\)
10: \(corp_R = corp_R \cup corp_{c}\)
11: train_NER(model_N, corp_N)
12: train_RE(model_R, corp_R)
13: \(i = i + 1\)
14: end while
15: \(K_f = \text{build_kg}(T_O, T_{conf})\)
16: return \(K_f\)

Algorithm 2 Constructing distantly-supervised corpus

Input: A part of text corpus \(D_i\), coarse-domain KG \(K_c\), new entities with high confidence \(E_{conf}\), new triplets with high confidence \(T_{conf}\), out-of-domain words \(W_O\)

Parameter: negative sample ratio \(ratio_n\), out-of-domain sample ratio \(ratio_o\)

Output: Distantly-supervision NER corpus \(corp_N\), distantly-supervision RE corpus \(corp_R\), overlapping entities \(E_O\), overlapping triplets \(T_O\)

1: \(\text{Let corp}_E = \{\}, \text{corp}_R = \{\}, E_O = \{\}, T_O = \{\}\)
2: sentence_num = len(\(D_i\))
3: \(j = 1\)
4: while \(j <= \text{sentence_num}\) do
5: \(\text{entities}_k, \text{entities}_c = \text{entity_matching}(D_i, K_c, E_{conf})\)
6: \(E_O = E_O \cup \text{entities}_k\)
7: \(corp_N = corp_N \cup \text{build_NER_sample}(D_i, \text{entities})\)
8: \(triplets_n, triplets_c = \text{entity_pair_matching}(D_i, K_c, T_{conf})\)
9: \(triplets = triplets_n \cup\) triplets_c
10: \(triplets_n = \text{get_negative_triplets}(D_i, \text{W}_O, \text{triplets}, ration_n)\)
11: \(corp_R = corp_R \cup \text{get_samples}(triplets_n)\)
12: \(corp_R = corp_R \cup \text{get_samples}(triplets)\)
13: \(T_O = T_O \cup triplets_k\)
14: \(j = j + 1\)
15: end while
16: return \(corp_N, corp_R, E_O, T_O\)

As shown in Algorithm 3, for discovering new entities \(E_{new}\), we will apply the trained \(model_N\) on dataset \(D_i\) and obtain \(entities\) that are disjoint with \(K_c\) (line 5 and 6). Then, we will merge \(entities\) with the previously-discovered entity set \(E_{new}\) (line 7). Finally, we will select the “high-confidence” entity as \(E_{conf}\) based on the mechanism above by the prediction probability and cumulative frequency (line 10). For the discovery of new triplets \(T_{new}\), we will enumerate entity pairs that are disjoint with the \(K_c\) (line 13-15). We then use the trained RE model and the predefined sample template to predict the relationship of the entity pairs and delete the triplets whose predicted relationship is NULL (line 16). Other processing is similar to the discovery of new entities.

After Algorithm 3, discovered entities specific to the fine domain are stored in \(E_{conf}\). Discovered triplets \(T_R\) (new relation, overlapping entity) and \(T_E\) (new relation, new entity) are stored in \(T_{conf}\). In the next iteration, Algorithm 2 will then use the updated \(E_{conf}\) and \(T_{conf}\) for building distantly-supervision corpus.

III. EXPERIMENTS

In this work, we used the adaptation of KG from the biomedical domain (coarse) to the oncology domain (fine) as an example to demonstrate the workflow of the KGDA framework, as well as to evaluate its effectiveness in practice.
Algorithm 3 Discovering fine-domain specific knowledge

**Input:** A part of text corpus text corpus $D_i$, coarse-domain KG $K_c$, new entities $E_{new}$, new entities with high confidence $E_{conf}$, new triplets $T_{new}$, new triplets with high confidence $T_{new}$. 

**Parameter:** NER model $model_N$, RE model $model_R$, probability threshold of the entity $th_pe$, frequency threshold of the entity $th_fe$, probability threshold of the triplet $th_pt$, frequency threshold of the triplet $th_ft$.

**Output:** $E_{new}$, $E_{conf}$, $T_{new}$, $T_{conf}$

1. Let $corp_E = \{\}$, $corp_R = \{\}$, $O = \{\}$; $O = \{\}$.
2. sentence_num = len($D_i$).
3. $j = 1$.
4. while $j <= $ sentence_num do
5.   $entities = NER\_prediction(D_i, model_N)$
6.   $entities = get\_new\_entities(\ entities, \ K_c)$
7.   $E_{new} = merge\_entity(\ E_{new}, \ entities)$
8.   $j = j + 1$
9. end while
10. $E_{conf} = get\_confidence\_entity(\ E_{new}, \ th_pe, \ th_fe)$
11. $j = 1$
12. while $j <= $ sentence_num do
13.   $entities = entity\_matching(\ D_i, \ K_c, \ E_{conf})$
14.   $pairs = enumerate\_pairs(\ entities)$
15.   $pairs = get\_new\_pairs(\ pairs, \ K_c)$
16. $triplets = RE\_prediction(D_i, \ pairs, \ model_R)$
17. $T_{new} = merge\_triplet(\ T_{new}, \ triplets)$
18. $j = j + 1$
19. end while
20. $T_{conf} = get\_confidence\_triplet(\ T_{new}, \ th_pt, \ th_ft)$
21. return $E_{new}$, $E_{conf}$, $T_{new}$, $T_{conf}$

A. Dataset

The dataset used in this paper is released by [7]. The source data is downloaded from 12 international journals in the oncology domain. PDF files of the papers were cleaned and converted to sentences. In total, we select 240,000 paragraphs as the unlabeled text corpus of the oncology domain $D$. The coarse-domain KG $K_c$ used in this work is the biomedical KG\(^1\), defines 18 entity types and 19 relationship types, including 5.2 million English entities and 7.34 million triplets.

B. Evaluation

Similar to the previous works [8], we evaluate our method in two schemes: held-out evaluation and manual evaluation. For the held-out evaluation, we reserved a part of the text corpus of $D$ as the test set. During the testing, we then compared the prediction results of the NER and RE models with the labels matched with $K_c$, and calculated the precision, recall, and F1 of the held-out dataset. Specifically, we use seqeval\(^2\) to evaluate the micro average precision, recall, F1 of NER. When evaluating the RE model, we perform relation classification prediction on the triplets existing in $K_c$ and corresponding entity pairs appearing in the held-out corpus. Finally, weighted average precision, recall, and F1 from the held-out evaluation will be reported.

As the labels of testing samples in the held-out evaluation are all inferred by distant supervision from the coarse domain, such scheme can only evaluate whether the trained model can capture the knowledge in the coarse domain, but cannot evaluate the ability of the models to discover new knowledge in the fine-domain. Therefore, we also adopted the manual evaluation scheme, consisting of the evaluations of: 1) the entities specific to fine domain $E_{conf}$, which are not presented in $K_c$; 2) the triplets of new relations $T_R$; 3) the triplets of new entities $T_E$. We randomly sampled 50 cases of $E_{conf}$, $T_R$, and $T_E$ respectively, then asked two physicians to manually label them for whether the entities and triplets are correct. As the number of named entities and triplets instances that are expressed in the corpus is unknown, we cannot estimate the recall of fine-domain KG. Therefore, we only show the precision of $E_{conf}$, $T_R$, and $T_E$. We fully recognize that the discovery of new knowledge in the fine-domain is an indispensable task for this work and we are recruiting more medical experts to conduct human reader study and performance evaluation for the proposed model.

C. Implementation settings

We divide the corpus $D$ into six equal subsets, and each subset contains around 40,000 sentences. We used $D_1$ to $D_5$ for model training and KG construction. We reserved $D_6$ for held-out evaluation. We tested BERT [9], BioClinicalBERT [10], biomed_RoBERTa [11] for initializing NER and RE models. Our experiments were run on an Ubuntu system computer with 4 NVIDIA A100 graphics cards. The learning rate, batch size, and epochs are set as 2E-05, 20, and 4, respectively. Hyperparameters $th_fe, th_pe, th_ft, th_pt$ are set as 2, 0.95, 3, and 0.97. The parameters $ratio_n$ and $ratio_o$ that control negative sampling are set to 0.2 and 0.3.

D. Held-out evaluation

| models          | precision | recall | F1   | precision | recall | F1   |
|-----------------|-----------|--------|------|-----------|--------|------|
| BERT            | 0.908     | 0.900  | 0.904| 0.987     | 0.949  | 0.967|
| BioClinicalBERT | 0.909     | 0.895  | 0.902| 0.988     | 0.957  | 0.972|
| biomed_RoBERTa | 0.908     | 0.901  | 0.905| 0.987     | 0.959  | 0.972|

Table I

Held-out evaluation of NER and RE model.

The results of the NER and RE models evaluated by the held-out dataset are shown in Table I, respectively. The KGDA frameworks initialized by the three pre-trained language models (BERT, BioClinicalBERT, and biomed_RoBERTa) all show good performance in held-out evaluations, demonstrating the robustness of our framework.

E. Manual evaluation

The number of all discovered entities ($E_O$), triplets ($T_O$), new entities with high confidence ($E_{conf}$), triplets representing

\(^1\)https://idea.edu.cn/bios.html

\(^2\)https://github.com/chakki-works/seqeval

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
new relations with overlapping entities (T_R), and triplets representing new relations with new entities (T_E) are shown in Table II, with each row belonging to one pre-trained language models used. Numbers of E_O and T_O have minor differences among different pre-trained language models, possibly due to the conflicts in strings matching of knowledge bases. We sampled 50 cases from E_conf, T_R, and T_E for manual evaluation, and the results are shown in Table III.

F. Ablation study

We investigated the impact of techniques employed by KGDA on its held-out experiment performance by removing the corresponding component from the framework. The results of the ablation analysis are shown in Table IV. Compared to the complete framework with w/o (cumulative), it can be seen that the using of accumulated data through iterations is beneficial for improving the generalization ability of NER and RE models. The held-out performances of the model without iteration indicates that the iterative training strategy can not only discover the specific knowledge in the fine domain but also maintain the ability to discover overlapping knowledge between the coarse and fine domain.

IV. Conclusion

In this paper, we propose an integrated, end-to-end framework for knowledge graph domain adaptation using distant supervision, which can be used to construct KG from fully unlabeled raw text data with the guidance of an existing KG. To deal with the potential challenges in distant supervision, which might limit the knowledge discovered from the new domain, we propose an iterative training strategy, which divides an unlabeled corpus into multiple corpora. For each new corpus to the model, we then combine the knowledge in the coarse domain with the knowledge identified from the previous corporuses for distantly-supervised training. By adopting the iterative training strategy, our proposed KGDA framework can discover not only knowledge that overlaps with the coarse domain, but also knowledge specific to the fine domain and unknown to the coarse domain, thus enabling coarse-to-fine domain adaptation. We implemented the adaptation from biomedical KG to the oncology domain in our experiments and verified the effectiveness of the KGDA framework through held-out and manual evaluation.

ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China (2022YFE0112200), the National Natural Science Foundation of China (U21A20520, 62325204), Science and Technology Project of Guangdong Province (2022A0505050014), the Key-Area Research and Development Program of Guangzhou City (202206030009).

REFERENCES

[1] Y. Zhang, M. Sheng, R. Zhou, Y. Wang, G. Han, H. Zhang, C. Xing, and J. Dong, “Hkgb: An inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians’ expertise incorporated,” Information Processing Management, vol. 57, no. 6, p. 102324, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0366453320308190
[2] D. A. Lindberg, B. L. Humphreys, and A. T. McCray, “The unified medical language system,” Yearbook of medical informatics, vol. 2, no. 01, pp. 41–51, 1993.
[3] M. Rotmensch, Y. Halpern, A. Timat, S. Horng, and D. Sontag, “Learning a health knowledge graph from electronic medical records,” Scientific reports, vol. 7, no. 1, pp. 1–11, 2017.
[4] N. Kertkeidkachorn and R. Ichise, “T2kg: An end-to-end system for creating knowledge graph from unstructured text,” in Workshops at the Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[5] M. Stewart and W. Liu, “Seq2kg: an end-to-end neural model for domain agnostic knowledge graph (not text graph) construction from text,” in Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning, vol. 17, no. 1, 2020, pp. 748–757.
[6] Q. Li, H. Li, H. Ji, W. Wang, J. Zheng, and F. Huang, “Joint bilingual name tagging for parallel corpora,” in Proceedings of the 21st ACM international conference on Information and knowledge management, 2012, pp. 1727–1731.
[7] S. Rezayi, H. Dai, Z. Liu, Z. Wu, A. Hehab, A. H. Burns, L. Zhao, D. Zhu, Q. Li, W. Liu et al., “Clinicalradibert: Knowledge-infused few shot learning for clinical notes named entity recognition,” in Machine Learning in Medical Imaging: 13th International Workshop, MLMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings. Springer, 2022, pp. 269–278.
[8] R. Zhou, W. Liu, and Y. Wang, “Agnostic knowledge graph (not text graph) construction from text,” in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009, pp. 1003–1011.
[9] J. D. M.-W. C. Kenton and L. K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of NAACL-HLT, 2019, pp. 4171–4186.
[10] E. Alsentzer, J. R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, W. Redmond, and M. B. McDermott, “Publicly available clinical bert embeddings,” NAACL HLT 2019, p. 72, 2019.
[11] S. Gururangan, A. Marasovic, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith, “Don’t stop pretraining: Adapt language models to domains and tasks,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8342–8360.