Leveraging Biomedical Ontologies to Boost Performance of BERT-Based Models for Answering Medical MCQs

Sahil Sahil[1,*], P Sreenivasa Kumar[1]

[1]Department of Computer Science & Engineering, Indian Institute of Technology Madras, Chennai

Abstract

Large-scale pretrained language models like BERT have shown promising results in various natural language processing tasks. However, these models do not benefit from the rich knowledge available in domain ontologies. In this work, we propose BioOntoBERT, a BERT-based model pretrained on multiple biomedical ontologies. We also introduce the Onto2Sen system to process various ontologies to generate lexical documents, such as entity names, synonyms and definitions, and concept relationship documents. We then incorporate these knowledge-rich documents during pretraining to enhance the model's "understanding" of the biomedical concepts. We evaluate our model on the MedMCQA dataset, a multiple-choice question-answering benchmark for the medical domain. Our experiments show that BioOntoBERT outperforms the baseline model BERT, SciBERT, BioBERT and PubMedBERT. BioOntoBERT achieves this performance improvement by incorporating only 158MB of ontology-generated data on top of the BERT model during pretraining, just 0.75% of data used in pretraining PubMedBERT. Our results demonstrate the effectiveness of incorporating biomedical ontologies in pretraining language models for the medical domain.

Keywords

Biomedical Ontologies, BERT, Medical Multiple Choice Question Answering

1. Introduction

Biomedical ontology research encompasses a variety of entities (from dictionaries of names for biological products to controlled vocabularies to principled knowledge structures) and processes (i.e., acquisition of ontological relations, integration of heterogeneous databases, use of ontologies for reasoning about biological knowledge) [1]. Biomedical ontologies include various aspects of medical terminologies such as symptoms, diagnosis and treatment.

Multiple-choice question-answering (MCQA) is a challenging task in general and in particular, in the domain of the medical field as the relevant knowledge is not commonly available in text corpora. The success of MCQA systems relies on striking a delicate balance between language understanding, domain-specific reasoning, and the incorporation of rich knowledge sources.

In the medical domain, the use of ontology-based QA systems has a very good potential to effectively capture domain-specific knowledge and provide accurate responses to medical

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil
*Corresponding author.
cs20s017@cse.iitm.ac.in (S. Sahil); psk@cse.iitm.ac.in (P. S. Kumar)
0009-0004-0167-5621 (S. Sahil); 0000-0003-2283-7728 (P. S. Kumar)

queries. By harnessing biomedical ontologies, these systems can depict intricate relationships among medical concepts, resulting in more precise and contextually aware answers.

Ontology-based multiple-choice question-answering systems are few in number, but Ontology-based QA systems have shown promise in capturing domain-specific knowledge and accurately answering medical questions [2] [3]. By leveraging biomedical ontologies, these systems can represent complex relationships between medical concepts, enabling more precise and contextually aware responses. A major limitation is that using these systems requires an understanding of the ontology structure in order to formulate queries. For example, queries may necessitate using intermediate concepts in the ontology when there is no direct relationship between the concepts in question.

Contextual word embedding models, such as BERT (Bidirectional Encoder Representations from Transformers) [4] have achieved state-of-the-art results in many NLP tasks. Initially tested in a general domain, models such as BioBERT[5], UmlsBERT [6], SciBERT [7], and PubmedBERT [8], have also been successfully applied in the biomedical domain by pretraining them on biomedical corpora. However, current biomedical applications of transformer-based NLP models do not incorporate structured expert domain knowledge from a biomedical ontology into their embedding pretraining process.

To illustrate the significance of biomedical ontology knowledge, let's consider a scenario where a medical question pertains to a specific rare disease. While a pretrained language model trained on a vast corpus may have encountered related terms or phrases, it may lack the medical domain-specific knowledge required to provide accurate and nuanced answers. In contrast, a biomedical ontology encompasses structured and domain-specific knowledge, including relationships, hierarchies, and semantic information about medical concepts. By integrating such ontology knowledge into our models, we can tap into a comprehensive and precise representation of medical domain knowledge, enabling more accurate and contextualized question-answering.

In light of this research gap, our study aims to bridge the divide between ontology-based approaches and deep learning models in the context of MCQA in the medical domain. Specifically, our objectives are:

- To overcome the challenges of ontology injection, including the computational overhead and annotation burden associated with large biomedical ontologies.
- To investigate techniques for integrating biomedical ontological knowledge with pretrained BERT models in MCQA systems.

In this paper, we present a novel approach that bridges the gap between ontology-based methods and pretrained language models, harnessing the strengths of both to enhance multiple-choice question-answering (MCQA) in the medical domain. Our contributions to this work can be summarized as follows:

1. Onto2Sen, a simple yet effective solution for Ontology Injection: We propose a unique solution called Onto2Sen system to generate a comprehensive ontology-backed sentence corpus, which serves as a valuable resource for enriching pretrained models with domain-specific knowledge. By incorporating this rich semantic information from biomedical domain ontologies into the models, we anticipate enhancing their contextual understanding and reasoning abilities.

2. Introducing BioOntoBERT: We propose BioOntoBERT, a pretrained BERT model that leverages various Biomedical Ontologies using the Onto2Sen generated corpus. BioOntoBERT surpasses several other biomedical BERT models, including PubmedBERT [8], SciBERT[7] and BioBERT [5], in terms of performance for multiple-choice question answering on the MedMCQA dataset.

Furthermore, BioOntoBERT demonstrates remarkable performance with just 158MB of pretraining data, significantly reducing the computational cost and carbon footprint associated with larger models. This aspect makes our novel approach not only effective but also environmentally friendly, addressing the growing concerns regarding energy consumption in deep learning models and highlighting the power of knowledge.

2. Related Work

Biomedical Multiple Choice Question Answering (MCQA) is a significant task in natural language processing. Various approaches have been proposed to improve the performance of MCQA systems by leveraging ontologies and pretrained language models.

As mentioned earlier, Ontology-based MCQA models are relatively limited, while Ontology-based question-answering systems have shown promise in capturing domain-specific knowledge and providing accurate answers to medical questions. For instance, in the case of XMQAS proposed by Midhunlal et al.[9], the system utilized natural language processing techniques and ontology-based analysis to process medical queries and extract relevant information from medical documents. Other approaches, like the one presented by Kwon et al.[10] for stroke-related knowledge retrieval, employed SPARQL templates and medical knowledge QA query ontology to transform queries into executable SPARQL queries for retrieving medical knowledge. However, these approaches have limitations due to their reliance on a template-based approach, which may restrict the flexibility and adaptability of the system.

In addition to ontology-based approaches, using pretrained models has significantly advanced MCQA systems. One notable example is PubmedBERT [8], a variant of BERT designed explicitly for biomedical text comprehension. These pretrained models, including PubmedBERT, have showcased remarkable performance in capturing medical terminologies and comprehending complex medical questions. Moreover, models like BioBERT [5], SciBERT[7], and UmlsBERT [6] have been finetuned for biomedical NLP tasks, exhibiting improved performance in various medical question-answering and information retrieval tasks. It is worth noting that these models are pretrained on extensive corpora, such as Pubmed abstracts entire medical dataset, which consists of over 3.1 billion words.

Less amount of work has been done in using external knowledge with neural networks in the biomedical multiple choice question answering domain, whereas in other domains like common sense reasoning several different approaches have been investigated for leveraging external knowledge sources. Sap et al.[11] introduce the ATOMIC graph with 877k textual descriptions of inferential knowledge (e.g. if-then relation) to answer causal questions. Lv et al.[12] propose to extract evidence from both structured knowledge bases such as ConceptNet and Wikipedia text and conduct graph-based representation and inference for commonsense reasoning.

He et al.[13] proposed a training procedure to infuse disease knowledge and augment pre-trained BERT models. Their experiments demonstrated improved performance in consumer health question answering, medical language inference, and disease name recognition. This motivates us to leverage the strengths of ontology which excel at representing complex medical concepts and terminologies. By integrating ontology and BERT-based models, we aim to enhance the capabilities of our MCQA system and improve its accuracy and effectiveness in addressing biomedical questions.

To bridge the gap between ontology-based approaches and deep learning models, the authors of [14] [15] [16] have explored techniques for ontology injection and infusing context. These approaches aim to enhance the models' language understanding and domain-specific reasoning capabilities by injecting ontological information into the models by modifying or adding new BERT layers or mapping the concepts and relationships of the ontology to the data. However, these models face various challenges in processing and incorporating large biomedical ontologies. The computational overhead required to handle and integrate the vast knowledge in such ontologies can be significantly high. Moreover, the process of mapping the ontology with the dataset and preparing annotated data demands substantial time and labour resources. The manual effort required for this task can be burdensome, hindering the scalability and practicality of these approaches.

3. Biomedical Ontologies

Biomedical ontologies play a critical role in the field of medicine by organizing and representing knowledge related to diseases, genes, anatomical structures, and medical concepts. They establish a standardized framework that captures and integrates information, promoting data sharing, interoperability, and knowledge discovery. We now briefly describe the prominent biomedical ontologies we use for our model:

1. **Disease Ontology (DO)** [17] (v1.2): The Disease Ontology is a standardized ontology created to offer the biomedical community consistent, reusable, and sustainable descriptions of human disease terms, phenotype characteristics, and related medical vocabulary disease concepts.
2. **Gene Ontology (GO)** [18] (v2023-04-01): It is a widely used ontology that focuses on representing the functional attributes of genes and gene products across different species. GO encompasses three main domains: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). BP describes biological processes in which genes are involved, MF represents the molecular functions they perform, and CC defines their cellular locations.
3. **Foundational Model of Anatomy Ontology (FMAO)** [19] (v5.0.0): FMAO is an ontology that aims to represent human anatomy in a detailed and structured manner. FMAO provides a hierarchical organization of anatomical structures, capturing spatial relationships and functional associations between different body parts.
4. **Precision Medicine Ontology** [20] (v4.0): It is a comprehensive ontology that represents medical concepts and their relationships in a standardized manner. Medicine Ontology covers various medical domains, including diseases, symptoms, treatments, diagnostic procedures, and medical devices.

Table 1
Different Biomedical Ontologies used

Ontology	Scope	Classes	# Object Properties	# Annotations	# subClass
FMAO Ontology	Anatomy	104721	139	51	262548
Bioassay Ontology	Pharmacology	904	17	34	981
Dental Ontology	Dentistry	2745	62	28	6507
Gene Ontology	Bioinformatics	84108	297	60	192606
Precision Medicine Ontology	Medicine	76155	95	23	122760
Disease Ontology	Pathology	11033	2	53	11063
Paediatrics Ontology	Paediatrics	1771	-	8	1760
HPS Ontology	Physiology	2920	86	34	3143
Mental Disease Ontology	Psychiatry	879	41	102	940

5. **Bioassay Ontology (BAO)** [21] (v1.1): The BAO focuses on establishing common reference metadata terms and definitions required for describing relevant information of low-and high-throughput drug and probe screening assays and results.
6. **Dental Ontology** [22] (v2016-06-27): It captures dental-related concepts and relationships, providing a standardized vocabulary for representing dental conditions, procedures, materials, and anatomical structures. It facilitates the integration of dental data and knowledge, supporting research, education, and clinical practice in dentistry.
7. **Pediatrics Ontology** (v2.0): Ontology focuses on representing pediatric healthcare-related concepts and their relationships. It covers various aspects of pediatric medicine, including diseases, developmental milestones, treatments, and interventions.
8. **Human Physiology Simulation Ontology (HPSO)** [23] (v1.1.1): HPSO captures the concepts and relationships related to the simulation and modelling of human physiology. It provides a standardized framework for representing physiological processes, organ interactions, and computational models.
9. **Mental Disease Ontology (MDO)** [24] (v2020-04-26): MDO represents mental disorders and related concepts. It offers a standardized vocabulary for categorizing and annotating mental diseases, symptoms, treatments, and diagnostic criteria.

4. Methodology

In this section, we present our approach for pretraining and fine-tuning a BERT[4] model on biomedical ontologies for multiple-choice question answering on the MedMCQA dataset. Our approach involves several key steps: data preparation, pretraining on biomedical ontologies, and fine-tuning the MedMCQA dataset. The code implementation is publicly available on GitHub[1].

4.1. Datasets

4.1.1. Multiple Choice Questions Dataset

We use the MedMCQA dataset[25], which consists of 1,94,000 multiple-choice questions on around 2400 healthcare topics and 21 medical subjects from one of the toughest entrance exams conducted for medical graduates in India, i.e., AIIMS and NEET PG. The diversity of questions

[1]https://github.com/sahillihas/BioOntoBERT

Table 2
Sample MCQA question from MedMCQA dataset with the correct answer as (A)

Question: Dentigerous cyst is likely to cause which neoplasia?	
(A) Ameloblastoma	(B) Adenocarcinoma
(C) Fibrosarcoma	(D) All of the above

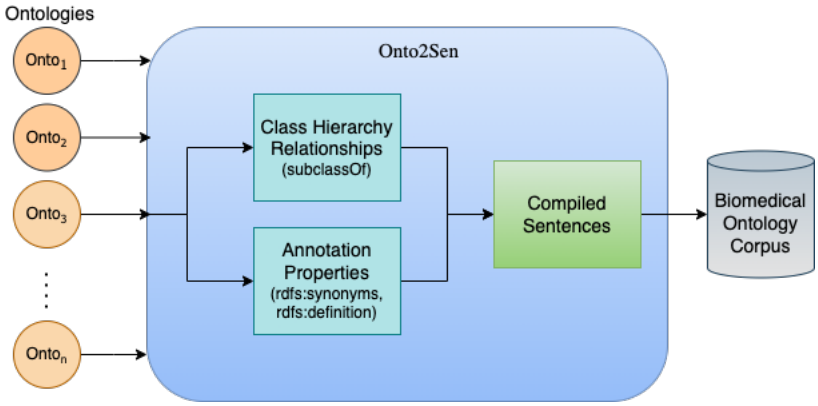

Figure 1: Proposed Onto2Sen Framework to generate BERT input corpus from the Ontologies

in the MedMCQA makes it a challenging dataset containing many aspects of medical knowledge; Table 2 illustrates one such example. Another distinguishing factor of this dataset is its questions are created for and by human experts. The dataset has three parts: the training set of 1,82,822 questions, the validation set of 4183 and the test set comprising 6150 questions, with an average token length of 12.35, 13.91 and 9.68, respectively. The answer choices are provided in the 'labels' column, encoded as integers 0, 1, 2, and 3. The ground truth for the test set is not publicly available. Hence we will be analysing the results on the validation set.

4.1.2. Ontology-based Sentence Generation

We propose a system called Onto2Sen to generate sentences from multiple ontologies curated from public resources mentioned in the previous section. It extracts concepts, annotations, and their properties from the ontology to form meaningful sentences. Onto2Sen preprocesses the ontologies and generates two types of sentences. The first type of sentence generated is from the subClass relationships. The second type of sentence is extracted from the relevant lexical annotation axioms in the ontology.
In the example shown in Figure 1, the Class Hierarchy Relationship sentences will contain the subClass property in the Disease Ontology (DO) allowing us to identify specific disease classifications. For instance, we can state that 'SPOAN syndrome is a neurodegenerative disease' using labels and identifiers in subClass relations. In addition, the transitive nature of the subclass properties is also utilized. Furthermore, Annotation Properties associated with diseases offer valuable insights into symptoms, synonyms and causal associations. For instance, we can describe that "SPOAN syndrome has synonym Spastic paraplegia" using the 'has_exact_synonym' annotation property.

We then used a natural language processing tool, spaCy, for preprocessing the compiled documents. We use these generated sentences as input to the model during pretraining to leverage the ontological knowledge.

After a study of the ontologies mentioned in Section 3, we find that using annotation properties and the class hierarchy for sentence generation is commonly applicable across all these ontologies and hence we adopt only these techniques for the present.

4.2. Pretraining Model

Pretraining is a crucial aspect of the BERT (Bidirectional Encoder Representations from Transformers) [4] model, which has revolutionized the field of natural language processing. In the context of BERT, pretraining refers to the initial phase where the model is trained on vast amounts of unlabeled text data, such as web documents or books. During this pretraining phase, BERT learns to generate contextualized representations of words and capture intricate semantic relationships by leveraging the bidirectional nature of transformers.

We propose a novel approach using Biomedical ontologies to pretrain the BERT model. As mentioned in the previous section, Onto2Sen can generate a corpus of meaningful sentences from different Biomedical ontologies. We use this generated corpus consisting of about 20M words which is a substantial volume of unlabeled text data related to the medical domain. The corpus was preprocessed and prepared for training, ensuring it was suitable for the subsequent steps.

The BERT model's pretraining phase involves two tasks: Masked Language Modelling (MLM) and Next Sentence Prediction. However, for our model, which incorporates biomedical ontologies, we focus on augmenting the Masked LM task and omit the Next Sentence Prediction task.

In the Masked LM task, we masked out 15 per cent of tokens in a sentence, and the model is trained to predict the original tokens given the context of the surrounding words. This approach will help the semantic understanding of medical terminologies by directly injecting biomedical ontology concepts and properties into the input sequence. As a result, the model will recognise and better understand medical concepts and terminologies effectively.

During the pretraining process, the BERT model was trained using the Adam optimizer, a widely adopted optimization algorithm for neural networks. The optimizer iteratively adjusted the model's parameters to minimize a predefined loss function, optimizing its ability to capture language patterns. Additionally, a learning rate scheduler was employed to dynamically adjust the learning rate at specific intervals, facilitating improved convergence and optimization of the model. The scheduler strategy, such as linear or exponential decay, was carefully selected based on experimentation and optimization.

These pretraining steps establish a well-built foundation for subsequent finetuning and proficient utilization of the BioOntoBERT model across diverse downstream natural language processing tasks.

4.3. Finetuning BERT

During the fine-tuning stage, we aim to train our BioOntoBERT model to accurately answer multiple-choice questions on the MedMCQA dataset without using any external context.

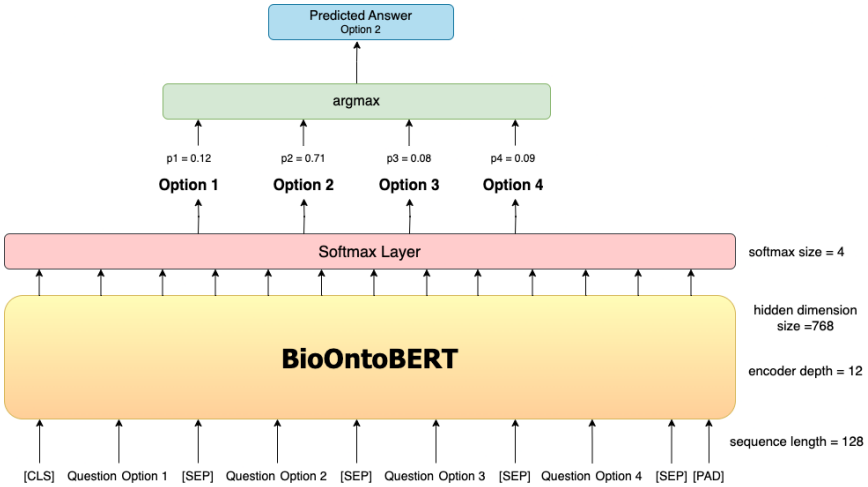

Figure 2: BioOntoBERT for multiple choice questions

Each multiple-choice question in the MedMCQA dataset was concatenated with its answer options to form a single input sequence of the form as shown in Figure 2.
Next, we performed tokenization on the dataset. Tokenization involves breaking down the questions and answers choices into smaller units called tokens, which the model can handle. This step ensures that the data is in a format suitable for the BioOntoBERT model to process. After the dataset is properly tokenized, we then train the BioOntoBERT model on this data.

During training, the model learned from the dataset by adjusting its internal parameters to better capture the relationships between questions and answer choices. The goal was to enhance the model's capacity to accurately choose the right answer when presented with a question. In this case, the labels were encoded in a one-hot format derived from integers. Throughout the training process, the model iteratively refined its understanding of the task by analyzing the patterns and context in the data. We carefully optimized the model's performance by adjusting various parameters, such as the learning rate and the number of training epochs.

Once the training was completed, we evaluated the performance of the finetuned BioOntoBERT model using the validation dataset. This evaluation allowed us to measure how well the model performed on unseen data and provided valuable insights into its ability to answer multiple-choice questions accurately.

During the fine-tuning process and subsequent evaluation of the BioOntoBERT model, a probability distribution is generated for each question's answer choices. The output probability distribution is denoted by $p1$, $p2$, $p3$ and $p4$ as shown in Figure 2. We identify the most likely answer choice by choosing the index associated with the highest probability.

5. Results

The main objective of this paper is to investigate the impact of incorporating biomedical ontology into the pretraining process of BERT models for the task of medical multiple-choice question answering. To achieve this objective, we developed a new pretrained model, BioOn-

Table 3
Accuracy and additional corpus size for different models on the MedMCQA dataset [25]. Statistics for prior BERT models are taken from their publications. [4] [5] [7] [8] .

Models	Corpus	Text Size	Accuracy
BERT	Wiki + Books	-	35%
BioBERT	PubMed	4.5B Words	38%
SciBERT	PMC + CS	3.2B words	39%
PubmedBERT	PubMed	3.1B words | 21GB	40%
BioOntoBERT (proposed)	Biomedical Ontologies	**20M words | 158 MB**	**42.72%**

toBERT, that is pretrained on a combination of 9 biomedical ontologies. We evaluated the performance of BioOntoBERT on the MedMCQA dataset, which contains a set of challenging medical questions curated by medical experts and compared it to the performance of other pretrained models, such as PubMedBERT[26], SciBERT[7] and BioBERT[5].

We conducted the pretraining of our BioOntoBERT model using the BERT base architecture, pretrained on English Wikipedia and BooksCorpus for 1M steps. BioOntoBERT was pretrained for 200K steps. The pretraining process involved a batch size of 32 and a learning rate scheduling of 5e-5. The pretraining and finetuning were both performed on a Tesla V100-PCIE-32GB GPU, with a maximum sequence length of 128. The pretraining of BioOntoBERT on ontology-generated sentences took approximately 10 hours only, whereas the pretraining times for PubmedBERT and BioBERT were reported as 5 days (120 hours) [8] and 10 days (240 hours) [5], respectively. For the finetuning process, a batch size of 32 and a learning rate of 1e-5 were selected. It took approximately 30 hours to complete the finetuning process due to the large size of the MedMCQA training data.

BioOntoBERT outperformed the baseline BERT-base, achieving a minimum accuracy of 42.72% in 10 runs. Furthermore, BioOntoBERT also outperformed PubMedBERT, which is pretrained on a huge corpus of biomedical text data. These results indicate that adding ontology data to the pretraining process can improve the performance of BERT models for medical question answering.

The comparison of models in Table 3 highlights the significance of the relatively small amount of additional ontology data we used to enhance the performance of our model. This finding suggests that the biomedical ontology we injected into the model is highly informative and beneficial, unlike much of the data in other corpora, which may be considered irrelevant.

During the evaluation, we also conducted a comparative analysis of the performance of BioOntoBERT, BERT, and PubmedBERT on various multiple-choice questions across different medical subjects. One evaluated question is in Table 2. Notably, BioOntoBERT correctly predicted the answer as (A) since the keywords 'Ameloblastoma', 'Adenocarcinoma', 'Fibrosarcoma' and 'Neoplasia' are present in the DOID ontology, BioOntoBERT model would have leveraged this knowledge. Whereas 'Dentigerous cyst' is not present in the DOID, Dentigerous cyst is a type of 'Odontogenic Cyst', and DOID contains a reference to 'Odontogenic Epithelium'. Odontogenic cysts and Odontogenic epithelium are closely related, as the former is derived from the remnants of the latter and forms as a result of abnormal developmental processes during tooth formation. In contrast, both BERT and PubmedBERT predicted the an-

Table 4
Subject-wise model comparison of PubMedBERT and BioOntoBERT on MedMCQA validation set of AIIMS MCQA. Statistics for PubMedBERT subject-wise are taken from [25]

Subject Name	PubMedBERT	BioOntoBERT	Ontology Used
Anatomy	39%	**41%**	✓
Biochemistry	49%	**50%**	✓
Dental	36%	**40%**	✓
ENT	**52%**	41%	✗
Medicine	47%	**48%**	✓
Microbiology	**44%**	40%	✗
Pathology	46%	**47%**	✓
Pharmacology	**46%**	42%	✓
Physiology	**56%**	54%	✓
Psychiatry	**56%**	50%	✓
Radiology	**31%**	28%	✗

swer as (D). This demonstrates an example instance of BioOntoBERT utilizing domain-specific knowledge.

The results presented in Table 4 demonstrate that BioOntoBERT exhibited superior performance compared to PubmedBERT across various subjects during pretraining, particularly when ontology data was available. Subjects like Anatomy, Biochemistry, Dental, Medicine, and Pathology showed notable improvements by including ontology data. However, for subjects such as ENT, Microbiology, and Radiology, where no ontology was used in our experiments, the benefits were not as evident. Additionally, for Pharmacology, Physiology and Psychiatry, the subject ontologies were not comprehensive enough to contribute significantly to question-answering capabilities. These findings underscore the significance of incorporating subject-specific ontology information to enhance the model's understanding and performance on domain-specific questions.

Importantly, we also evaluated the impact of the size and complexity of ontologies on the performance of the models. Surprisingly, we observed that the size or the number of concepts and properties in the ontologies did not necessarily correlate with improved question-answering performance. This suggests that the relevance and quality of the ontology data are crucial factors in enhancing the model's understanding and reasoning capabilities rather than the sheer quantity of information.

6. Conclusions

This study introduces the Onto2Sen system, which incorporates annotation-based and class-hierarchical sentences from ontologies to enhance the performance of a language model. It is the first instance of leveraging such knowledge in pretraining a language model for biomedical natural language processing tasks. The BioOntoBERT model, pretrained on biomedical ontologies, outperforms other models, including PubMedBERT, in multiple-choice question-answering tasks within the medical domain, effectively capturing medical terminologies. By achieving improved results with just 158MB of pretraining data, our approach not only enhances performance but also significantly reduces computational costs, making it a more sustainable approach to model training.

7. Future work

Firstly, the selection and incorporation of appropriate biomedical ontologies remain an ongoing challenge. While we employed several ontologies in our pretraining process, there are numerous other ontologies available that could potentially contribute to even better performance. Secondly, although BioOntoBERT exhibits impressive proficiency in language understanding and representation, it lacks advanced reasoning capabilities on ontologies. The model primarily captures contextual relationships and semantic information but does not possess explicit reasoning mechanisms to infer complex logical connections within ontologies. This limitation suggests avenues for future research, focusing on incorporating reasoning abilities into language models trained on biomedical ontologies.

References

[1] O. Bodenreider, A. Burgun, Biomedical ontologies, Medical Informatics: Knowledge Management and Data Mining in Biomedicine (2005) 211–236.

[2] Q. Guo, M. Zhang, Question answering based on pervasive agent ontology and semantic web, Knowledge-Based Systems 22 (2009) 443–448.

[3] A. Arbaaeen, A. Shah, Ontology-based approach to semantically enhanced question answering for closed domain: A review, Information 12 (2021) 200.

[4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[5] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, Biobert: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics 36 (2020) 1234–1240.

[6] G. Michalopoulos, Y. Wang, H. Kaka, H. Chen, A. Wong, Umlsbert: Clinical domain knowledge augmentation of contextual embeddings using the unified medical language system metathesaurus, arXiv preprint arXiv:2010.10391 (2020).

[7] I. Beltagy, K. Lo, A. Cohan, Scibert: A pretrained language model for scientific text, arXiv preprint arXiv:1903.10676 (2019).

[8] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, Domain-specific language model pretraining for biomedical natural language processing, ACM Transactions on Computing for Healthcare (HEALTH) 3 (2021) 1–23.

[9] M. Midhunlal, M. Gopika, Xmqas-an ontology based medical question answering system, International Journal of Advanced Research in Computer and Communication Engineering 5 (2016) 929–932.

[10] S. Kwon, J. Yu, S. Park, J.-A. Jun, C.-S. Pyo, Stroke medical ontology qa system for processing medical queries in natural language form, in: 2021 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2021, pp. 1649–1654.

[11] M. Sap, R. Le Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, Y. Choi, Atomic: An atlas of machine commonsense for if-then reasoning, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 3027–3035.

[12] S. Lv, D. Guo, J. Xu, D. Tang, N. Duan, M. Gong, L. Shou, D. Jiang, G. Cao, S. Hu, Graph-

based reasoning over heterogeneous external knowledge for commonsense question answering, in: Proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 8449–8456.
[13] Y. He, Z. Zhu, Y. Zhang, Q. Chen, J. Caverlee, Infusing disease knowledge into bert for health question answering, medical inference and disease name recognition, arXiv preprint arXiv:2010.03746 (2020).
[14] T. R. Goodwin, D. Demner-Fushman, Enhancing question answering by injecting ontological knowledge through regularization, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing, volume 2020, NIH Public Access, 2020, p. 56.
[15] L. He, S. Zheng, T. Yang, F. Zhang, Klmo: Knowledge graph enhanced pretrained language model with fine-grained relationships, in: Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 4536–4542.
[16] K. Faldu, A. Sheth, P. Kikani, H. Akbari, Ki-bert: Infusing knowledge context for better language and domain understanding, arXiv preprint arXiv:2104.08145 (2021).
[17] L. M. Schriml, C. Arze, S. Nadendla, Y.-W. W. Chang, M. Mazaitis, V. Felix, G. Feng, W. A. Kibbe, Disease ontology: a backbone for disease semantic integration, Nucleic acids research 40 (2012) D940–D946.
[18] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al., Gene ontology: tool for the unification of biology, Nature genetics 25 (2000) 25–29.
[19] C. Rosse, J. L. Mejino Jr, The foundational model of anatomy ontology, in: Anatomy ontologies for bioinformatics: principles and practice, Springer, 2008, pp. 59–117.
[20] L. Hou, M. Wu, H. Y. Kang, S. Zheng, L. Shen, Q. Qian, J. Li, Pmo: A knowledge representation model towards precision medicine, Math. Biosci. Eng 17 (2020) 4098–4114.
[21] U. Visser, S. Abeyruwan, U. Vempati, R. P. Smith, V. Lemmon, S. C. Schürer, Bioassay ontology (bao): a semantic description of bioassays and high-throughput screening results, BMC bioinformatics 12 (2011) 1–16.
[22] W. D. Duncan, T. Thyvalikakath, M. Haendel, C. Torniai, P. Hernandez, M. Song, A. Acharya, D. J. Caplan, T. Schleyer, A. Ruttenberg, Structuring, reuse and analysis of electronic dental data using the oral health and disease ontology, Journal of Biomedical Semantics 11 (2020) 1–19.
[23] M. Gündel, E. Younesi, A. Malhotra, J. Wang, H. Li, B. Zhang, B. de Bono, H.-T. Mevissen, M. Hofmann-Apitius, Hupson: the human physiology simulation ontology, Journal of biomedical semantics 4 (2013) 1–9.
[24] J. Hastings, W. Ceusters, M. Jensen, K. Mulligan, B. Smith, Representing mental functioning: Ontologies for mental health and disease (2012).
[25] A. Pal, L. K. Umapathi, M. Sankarasubbu, Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering, in: Conference on Health, Inference, and Learning, PMLR, 2022, pp. 248–260.
[26] Y. Peng, S. Yan, Z. Lu, Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets, arXiv preprint arXiv:1906.05474 (2019).