Entity Coreference and Co-occurrence Aware Argument Mining from Biomedical Literature

Boyang Liu[1], **Viktor Schlegel**[1,2], **Riza Batista-Navarro**[1], **Sophia Ananiadou**[1]
[1] Department of Computer Science, The University of Manchester, UK
[2] ASUS Intelligent Cloud Services (AICS), Singapore
{boyang.liu-2@postgrad., riza.batista@, sophia.ananiadou@}manchester.ac.uk
viktor_schlegel@asus.com

Abstract

Biomedical argument mining (BAM) aims at automatically identifying the argumentative structure in biomedical texts. However, identifying and classifying argumentative relations (AR) between argumentative components (AC) is challenging since it not only needs to understand the semantics of ACs but also need to capture the interactions between them. We argue that entities can serve as bridges that connect different ACs since entities and their mentions convey significant semantic information in biomedical argumentation. For example, it is common that related AC pairs share a common entity. Capturing such entity information can be beneficial for the Relation Identification (RI) task. In order to incorporate this entity information into BAM, we propose an Entity Coreference and Co-occurrence aware Argument Mining (ECCAM) framework based on an edge-oriented graph model for BAM. We evaluate our model on a benchmark dataset and from the experimental results we find that our method improves upon state-of-the-art methods.

1 Introduction

There is a growing interest in evidence-based decision making in the biomedical field, as it can assist medical practitioners in selecting the best treatment for a given medical case. However, extracting relevant evidence from vast amounts of biomedical publications is time-consuming for practitioners. Thus, biomedical Argument Mining (BAM), which is the application of Argument Mining (AM) to biomedical texts, is proposed to automatically extract argumentative structures in biomedical texts by identifying Argument Components (AC) and Argument Relations (AR) between ACs (Mayer et al., 2020). BAM includes three primary tasks (Si et al., 2022): (1) argument component identification (ACI)—i.e., distinguishing argumentative components from non-argumentative content; (2) argument component classification (ACC)—categorizing ACs into different types (e.g., claim, and evidence); and (3) relation identification (RI)—recognizing ARs (e.g., support, attack, or none) between a pair of ACs.

Among these tasks, the RI task is the hardest one and existing models tend to underperform on this task, compared to the ACI and ACC tasks (Mayer et al., 2020; Galassi et al., 2021; Si et al., 2022). One possible reason is that these models do not incorporate the information about the co-occurrence of common entities between different ACs. This is a valuable source of semantic information and can be particularly important in BAM. As shown in Fig 1, the AC pairs connected by an AR share entities with coreference relations. Furthermore, entity co-occurrence suggests the direction of the ARs (i.e. ACs with several entities are usually the tail of ARs, like in Figure 1).

Based on this intuition, we propose an Entity Coreference and Co-occurrence aware Argument Mining (ECCAM) framework that effectively captures ARs through entity coreference and co-occurrence. ECCAM is a graph-based model. We build a heterogeneous graph that consists of nodes that represent ACs and entities, and edges between nodes. The entity nodes can serve as bridges that connect different ACs. Considering that the entity coreference and co-occurrence relations exist between nodes and thus are represented by edge embeddings, we employ an edge-oriented graph model (Christopoulou et al., 2019) that learns edge representations of any two connected nodes by combing all paths between the two nodes. This enables information flow between different relations and iteratively updates the edge representations, which is finally used as the representations of ARs. Here, the edges between AC nodes and entity nodes are used to pass the entity coreference information while the edges between entity nodes aim to leverage entity co-occurrence information.

Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023), pages 54–60
July 13-14, 2023 ©2023 Association for Computational Linguistics

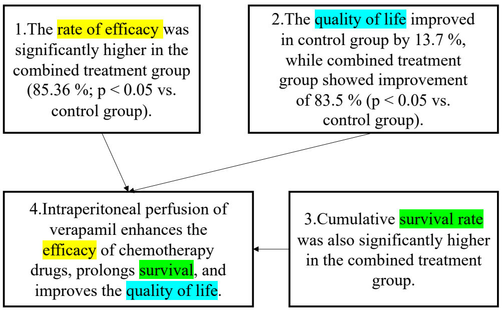

Figure 1: Part of the argument structure of the abstracts from PubMed 23589316. Each text in the rectangle represents an AC. Two ACs connected with an arrow means that there is an AR between them. Entities with the same colour are in the same coreference clusters.

Our contributions are shown below:

- To our best knowledge, this paper is the first to incorporate entity coreference and co-occurrence information into an argument mining model.
- We propose the ECCAM framework based on an edge-oriented model to leverage the entity coreference and co-occurrence information.
- Experimental results show that the entity coreference and co-occurrence information can improve the performance of the RI task significantly.

2 Related Work

Recently, the research community has shown growing interest in the task of BAM. Mayer et al. (2018) created a dataset by annotating ACs within randomized controlled trial abstracts and employ the SubSet Tree Kernel to classify the types of ACs with Bag-Of-Words of biomedical text as input. Further, a dataset with both ACs and ARs are created by Mayer et al. (2020) to deal with three tasks of BAM: ACI, ACC and RI. Various contextualized word embeddings, such as BERT (Devlin et al., 2019), BioBERT (Lee et al., 2020), SciBERT (Beltagy et al., 2019a), and RoBERTa (Liu et al., 2019) are explored to address these tasks. Liu et al. (2022) incorporate zoning information (such as background, result and conclusion) to tackle the ACI and ACC tasks at the same time. Galassi et al. (2021) employed a multi-task framework with an attentive residual network to address the ACC task, RI task, and link prediction task of BAM, based on the assumption that ACs had been detected. SeqMT (Si et al., 2022) also assumes the ACI task is solved and pays attention to the ACC and RI tasks. It utilises a multi-task learning approach to benefit from the sequential dependency between the ACC and RI tasks by transferring the representation of the input and output of the ACC task to the RI task.

However, none of the previous models leverages the entity coreference and co-occurence information for the RI task, which is the focus of this paper.

3 Model Architecture

Following previous models (Galassi et al., 2018; Si et al., 2022; Galassi et al., 2021), we assume that the outputs of the ACI task are provided, i.e., all ACs have been detected without AC types. Inspired by Christopoulou et al. (2019) who employ an edge-oriented graph model (Christopoulou et al., 2018) to leverage interactions among sentences that share the same entities for document-level relation extraction, we propose a framework for the RI task based on a similar edge-oriented model. Our framework contains three parts: the entity cluster extraction module, the document encoder module and the entity co-occurrence and coreference-aware argument mining module.

3.1 Entity Cluster Extraction Module

Since we cannot assume golden annotation of entities and their mentions, the first step of our framework is to identify all named entities in the AM dataset. One simple way is to train a model on biomedical coreference resolution datasets and use such model to predict entity clusters. However, most biomedical coreference resolution datasets concern diseases (Doğan et al., 2014; Li et al., 2015), species (Gerner et al., 2010; Pafilis et al., 2013) or proteins and genes (Wei et al., 2015; Collier and Kim, 2004). Most noteably, datasets with entity annotations related to cancer experiments, such as "quality of life" and "survival rate", are absent from the literature, while the biomedical argument mining dataset (Mayer et al., 2020) is about the cancer research. Thus, we choose another method that first predicts the entities in the AM datasets, and then disambiguates them by mapping to the Unified Medical Language System (UMLS) (Bodenreider, 2004) to obtain a unique identifier of a medical concept to obtain entity clusters.

Specifically, we use the Transformed NER model (Stylianou and Vlahavas, 2021) to obtain entities in the abstracts. This model is trained on the EMB-NLP (Nye et al., 2018) dataset and thus extracts four types of entities, namely Patient, Intervention, Comparison and Outcome (PICO). Since ACs mainly exist in sentences that describe experimental results and conclusions (Liu et al., 2022), we only use outcome entities, to avoid noise. Then, we use all the extracted entities as input of the QuickUMLS tool (Soldaini and Goharian, 2016) to obtain the corresponding UMLS identifiers (IDs). It is worth mentioning that given one mention, the QuickUMLS usually returns multiple IDs. We handle this situation as follows: when there is an exact match between the predicted entities and the given mention, we will use the ID of the exactly matched entity as the ID for the mention. If there is no exact match entity, we will reserve all the entities whose Jaccard similarity score with the given mention is higher than a specific threshold.

All entities in a document that share the same UMLS ID in a document form an entity cluster. The extracted entity clusters are denoted as $C = \{C_1, C_2, ..., C_{l_c}\}$, where $C_i = \{m_1, m_2, ..., m_l\}$, and l is total number of mentions in the i-th cluster.

3.2 Document Encoder Module

Given a document $D = \{t_1, t_2, ..., t_n\}$ consisting of n tokens as the input of our framework, a SciBERT model is employed as the encoder to generate the embeddings of tokens $X = \{x_1, x_2, ..., x_n\}$ in D.

$$X = SciBERT(D) \tag{1}$$

To leverage the entity coreference and co-occurrence information, a two-step method is proposed to generate the embeddings of all entities occurring in the document D. First, our model generates the embeddings of each mention m_i by averaging all token embeddings $\{x_{i1}, x_{i2}, ..., x_{iM}\}$ in m_i. Similarly, the embeddings of each AC n_a and each entity n_e are also an average of the tokens in the AC and entity, respectively.

3.3 Graph Construction

We initially construct a heterogeneous graph that consists of two different types of nodes (AC nodes and entity nodes) and three types of edges between the nodes. The rules for edge generation are outlined below.

AC-AC edge. If two ACs are adjacent, an edge will connect the two AC nodes. There are two situations where two ACs are adjacent. The first one is that the sequences of this two ACs are adjacent in the document. Another situation is that the words between the two ACs in the document are all non-argumentative. This type of edge is used to learn the context.

AC-Entity edge. If an AC mentions an entity at least once, there will be an edge between the entity and the related AC. This type of edge is used to learn the coreference information.

Entity-Entity edge. We connect all entity pairs so that the model can learn which co-occurrence of entity pairs is helpful for the RI task.

We use a concatenation operation to get the representation of an edge $e_{ij} = [n_i, n_j]$ given the representations of the source node n_i and the destination node n_j of the edge, where $n_i, n_j \in n_e \cup n_a$.

3.4 Entity Co-occurrence and Coreference aware Argument Mining model Module

Given the constructed graph, we employ an edge-oriented graph model (Christopoulou et al., 2018) to leverage the entity coreference and co-occurrence information. The model uses a two step method to iteratively update the edge embeddings of two nodes based on the paths between the two nodes.

First, a path between two nodes i and j is generated using intermediate nodes k. Then, the representations of two consecutive edges e_{ik} and e_{kj} are combined by a modified bilinear transformation. Through this action, an edge representation of double the length is generated. All existing paths between i and j through k are combined. The i, j, and k nodes can be either entity nodes or AC nodes. Intermediate nodes without adjacent edges to the target nodes are ignored. Formally, this is written as:

$$f(e_{ik}^{(l)}, e_{kj}^{(l)}) = \sigma(e_{ik}^{(l)} \otimes W e_{kj}^{(l)}) \tag{2}$$

where σ is the sigmoid non-linear function, W is a learned parameter matrix, $\otimes$ represents element-wise multiplication, l denotes the length of the edge and e_{ik} refers to the representation of the edge between nodes i and k.

At the second step, the original (short) edge representation and the new (longer) edge representation resulting from Equation 2 is aggregated as follows:

$$e_{ij}^{(2l)} = \beta e_{ij}^{(l)} + (1-\beta) \sum_{k \neq i,j} (e_{ik}^{(l)}, e_{kj}^{(l)}) \tag{3}$$

where $\beta \in [0, 1]$ is a scalar used to assign the weight of the shorter edge representation.

A finite number N of iterations is conducted for the two steps. The final length of path is directly proportional to the number of iterations. After N iterations, the number of edges of the longest path will be up to 2^N.

3.5 Classification Module

Finally, to classify relations between AC pairs, we incorporate a softmax classifier which takes the AC-to-AC edges e_{aa} as input:

$$y = softmax(We_{aa} + b) \tag{4}$$

where W and b are learned parameters of the classification layer. The whole model is trained end-to-end by minimising the cross-entropy loss between predicted and gold ACs.

4 Experiment

4.1 Datasets

Following Si et al. (2022), we use the AbstRCT (Mayer et al., 2020) benchmark to evaluate our model and compare it with previous approaches. The AbstRCT dataset is composed of three categories of ACs (major claim, claim, and evidence) and two kinds of ARs (support and attack). It consists of three parts, with the largest being the neoplasm corpus, which is divided into training, development, and testing sets. Moreover, there are two additional test sets. The first one solely consists of abstracts related to glaucoma, while the second one is a mixed set containing 20 abstracts for each disease in the dataset (neoplasm, glaucoma, hypertension, hepatitis, and diabetes).

	Documents	All ARs	Avg. AR
Neo_train	350	1427	4.1
Neo_dev	50	210	4.2
Neo_test	100	424	4.2
Gla_test	100	367	3.7
Mix_test	100	329	3.3

Table 1: Statistics of the AbstRCT dataset. The data statistics of the three test sets are reported separately. Here, *Neo*, *Gla* and *Mix* represent neoplasm, glaucoma and mixed, respectively.

4.2 Implementation

We use the same train-development-test split for the AbstRCT dataset as was used in Si et al. (2022). We fine-tune cased SciBERT (Beltagy et al., 2019b) and set the maximum sequence length to 256. A learning rate of $2 \cdot 10^{-5}$ is used. We train for 50 epochs with early stopping to avoid overfitting. Our model is implemented in PyTorch (Paszke et al., 2019). We employ an AdamW optimizer (Loshchilov and Hutter, 2019) for parameter optimization and report the macro-averaged F1 scores of models trained with three different random seeds.

4.3 Baselines

In order to evaluate our proposed method, we compare it with the following baselines:

ResArg (Galassi et al., 2018) is a hybrid of residual networks and long short-term memory network. This model is designed to tackle both the ACC and RI tasks simultaneously.

ResAttArg (Galassi et al., 2021) is an upgraded version of ResArg model featuring an attention module. ResArg and ResAttArg have two versions: an average version that calculates the final scores as an average of scores from 10 distinct networks trained with 10 different seeds, and an ensemble version that assigns the class based on the majority vote of the same 10 networks.

SeqMT (Si et al., 2022) utilises a multi-task learning approach to benefit from the sequential dependency between the ACC and RI tasks. It transfers the representation of the input and output of the ACC task to the RI task.

4.4 Main Results

We report the main results in Table 2. It can be observed that our model improves upon the state-of-the-art on two of three test sets even though our model is a single task model while all other baselines are multi-task/ensemble models. To be specific, our model outperforms the current best model on the neoplasm test set by 1.68% F1 score and the mixed test set by 1.11% F1 score. However, there is a gap between the performance of our model and SeqMT on the glaucoma test set. This might be due to the lack of multi-task training: compared with the results of SeqMT, the performance of the single task version model of SeqMT(SeqMT(-$\mathcal{L}_{acc}$)) similarly experiences a large drop of performance of 8.44% F1 score. Without the additional signal

models	NEO	GLA	MIX
RA(avg)	59.15	57.23	60.31
RA(Ensemble)	63.16	61.86	68.35
RAA(avg)	66.49	62.68	63.47
RAA(Ensemble)	70.92	68.40	67.66
SeqMT(-$\mathcal{L}_{acc}$)	68.58	64.83	70.30
SeqMT	71.24	**73.27**	72.71
ECCAM	**72.92**	68.96	**73.82**

Table 2: Main results of different models. The best scores are marked in bold. All the results of baselines are copied from the related papers. Here, *NEO*, *GLA* and *MIX* represent neoplasm, glaucoma and mixed.

from the ACC task, SeqMT(-$\mathcal{L}_{acc}$) performs significantly worse than our model on the glaucoma test set by 4.13% F1 score.

4.5 Ablation Study

To validate the effects of the entity coreference and entity co-occurrence information, we conduct two ablation experiments. **ECCAM(-EE)** is a model where the edges between the entities are excluded to test whether the entity co-occurrence information can improve the performance of RI. **ECCAM(-EA)** aims to reveal the impact of entity coreference information by removing both edges between entities and between entities and ACs. The results in Table 3 show the effectiveness of the entity coreference and entity co-occurrence information. Without the entity co-occurrence information, the performance of our model drops by 0.84%, 2.52% and 1.43% F1 score on the neoplasm, glaucoma and mixed test sets, respectively. The performance of **ECCAM(-EA)** decreases even more significantly—2.3%, 5.12% and 4.09% F1 score on the neoplasm, glaucoma and mixed test sets—showing the positive impact of entity coreference information.

models	NEO	GLA	MIX
ECCAM	**72.92**	**68.96**	**73.82**
ECCAM(-EE)	72.08	66.44	72.39
ECCAM(-AE)	70.62	63.84	69.73

Table 3: Ablation study of our model. **ECCAM(-EE)** is a model where the edges between the entities are excluded. **ECCAM(-EA)** removes both edges between entities and between entities and ACs. The best scores are marked in bold.

iterations	DEV	NEO	GLA	MIX
$N = 1$	60.20	65.32	53.49	66.90
$N = 2$	62.75	67.34	58.67	71.14
$N = 3$	66.04	71.84	63.37	71.80
$N = 4$	67.31	72.92	**68.96**	**73.82**
$N = 5$	**69.64**	**73.85**	66.30	68.83

Table 4: Results of hyper-parameter analysis. Here, *NEO*, *GLA* and *MIX* represent the results of on the neoplasm, glaucoma and mixed test sets, respectively. *DEV* denotes the results on the development set.

4.6 Hyper-parameter Analysis

We further test, whether number of iterations N affects the model performance on the three different test sets. We conduct experiments with $N = 1, 2, 3, 4, 5$. The results are shown in Table 4. From the results on the development set we can see that as the number of iterations increases, the performance of the model on the development set also increases. However, though our model obtains the best score on the neoplasm test set when $N = 5$, considering all three test sets, the best overall performance is achieved with four iterations. It is worth noting that the abstracts in the development set are all about neoplasm. Taking all these results into consideration, we can conclude that the more iterations the edges representations are updated, the more information is utilised from more distant nodes, with too many iterations causing overfitting.

5 Conclusion

In this paper, we propose the ECCAM model that is based on an edge-oriented graph model (Christopoulou et al., 2019) to incorporate entity coreference and co-occurrence information into BAM. We introduce edges between entity nodes and AC nodes in a heterogeneous graph to help our model capture entity coreference and co-occurrence information respectively. Experiments on the AbstRCT dataset show the effectiveness of these two types of information for the RI task. In the future, we will apply our method to other argument mining domains, such as student essays (Eger et al., 2017).

Limitations

Although our model improves upon state-of-the-art methods of BAM by incorporating entity coreference and co-occurrence information, there are still some limitations to our model. First, it is not easy to apply our model to other domains where no coreference resolution tool is available. Second, the number of nodes and edges of the generated heterogeneous graph will become enormous if the documents are long and many entities are extracted, which requires more GPU resources.

References

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019a. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China. Association for Computational Linguistics.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019b. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270.

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2018. A walk-based model on entity graphs for relation extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 81–88, Melbourne, Australia. Association for Computational Linguistics.

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4925–4936, Hong Kong, China. Association for Computational Linguistics.

Nigel Collier and Jin-Dong Kim. 2004. Introduction to the bio-entity recognition task at jnlpba. In *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP)*, pages 73–78.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Rezarta Islamaj Doğan, Robert Leaman, and Zhiyong Lu. 2014. Ncbi disease corpus: a resource for disease name recognition and concept normalization. *Journal of biomedical informatics*, 47:1–10.

Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural end-to-end learning for computational argumentation mining. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11–22, Vancouver, Canada. Association for Computational Linguistics.

Andrea Galassi, Marco Lippi, and Paolo Torroni. 2018. Argumentative link prediction using residual networks and multi-objective learning. In *Proceedings of the 5th Workshop on Argument Mining*, pages 1–10, Brussels, Belgium. Association for Computational Linguistics.

Andrea Galassi, Marco Lippi, and Paolo Torroni. 2021. Multi-task attentive residual networks for argument mining. *arXiv preprint arXiv:2102.12227*.

Martin Gerner, Goran Nenadic, and Casey M Bergman. 2010. Linnaeus: a species name identification system for biomedical literature. *BMC bioinformatics*, 11(1):1–17.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Jiao Li, Yueping Sun, R Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Thomas C Wiegers, and Zhiyong Lu. 2015. Annotating chemicals, diseases, and their interactions in biomedical literature. In *Proceedings of the fifth BioCreative challenge evaluation workshop*, pages 173–182. The Fifth BioCreative Organizing Committee.

Boyang Liu, Viktor Schlegel, Riza Batista-Navarro, and Sophia Ananiadou. 2022. Incorporating zoning information into argument mining from biomedical literature. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6162–6169, Marseille, France. European Language Resources Association.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Tobias Mayer, Elena Cabrio, Marco Lippi, Paolo Torroni, Serena Villata, et al. 2018. Argument mining on clinical trials. In *COMMA*, pages 137–148.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based argument mining for healthcare applications. In *ECAI 2020*, pages 2108–2115. IOS Press.

Benjamin Nye, Junyi Jessy Li, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace. 2018. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 197–207, Melbourne, Australia. Association for Computational Linguistics.

Evangelos Pafilis, Sune P Frankild, Lucia Fanini, Sarah Faulwetter, Christina Pavloudi, Aikaterini Vasileiadou, Christos Arvanitidis, and Lars Juhl Jensen. 2013. The species and organisms resources for fast and accurate identification of taxonomic names in text. *PloS one*, 8(6):e65390.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Jiasheng Si, Liu Sun, Deyu Zhou, Jie Ren, and Lin Li. 2022. Biomedical argument mining based on sequential multi-task learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1.

Luca Soldaini and Nazli Goharian. 2016. Quickumls: a fast, unsupervised approach for medical concept extraction. In *MedIR workshop, sigir*, pages 1–4.

Nikolaos Stylianou and Ioannis Vlahavas. 2021. Transformed: End-to-end transformers for evidence-based medicine and argument mining in medical literature. *Journal of Biomedical Informatics*, 117:103767.

Chih-Hsuan Wei, Hung-Yu Kao, and Zhiyong Lu. 2015. Gnormplus: an integrative approach for tagging genes, gene families, and protein domains. *BioMed research international*, 2015.