Exploring Pair-Aware Triangular Attention for Biomedical Relation Extraction

Lei Chen*
Department of Computer Science,
The University of Hong Kong
Hong Kong, China
lchen@cs.hku.hk

Junhao Su
Department of Computer Science,
The University of Hong Kong
Hong Kong, China
jhsu@cs.hku.hk

Tak-Wah Lam
Department of Computer Science,
The University of Hong Kong
Hong Kong, China
twlam@cs.hku.hk

Ruibang Luo*
Department of Computer Science,
The University of Hong Kong
Hong Kong, China
rbluo@cs.hku.hk

ABSTRACT

Biomedical relation extraction (BioRE) has become a research hotspot recently due to its crucial role in facilitating clinical diagnosis, treatment, and medical discovery. The advent of domain-specific language models, such as BioBERT and PubMedBERT customized for the biomedical domain, has revolutionized this task by fully learning contextualized entity representations and achieving remarkable performance. However, we argue that relying solely on entity-level modeling while neglecting pair-aware representations can lead to sub-optimal results, particularly in the complicated context of the biomedical literature. To address this issue, in this paper, we propose a novel **Tri**angular **A**ttention framework for **Bio**medical **R**elation **E**xtraction (called TriA-BioRE) to comprehensively capture pair-aware representations in the biomedical domain. Specifically, we present a triangular attention module, including two triangular multiplications utilizing outgoing and incoming edges, and two triangular self-attention operations centered on the starting and ending nodes, respectively, together to enhance the pair-level modeling omnidirectionally for better BioRE performance. Extensive experiments on three biomedical datasets demonstrate that TriA-BioRE achieves substantially better results than its strong competitors in BioRE task. For reproducibility, our code and data are available at https://github.com/JasonCLEI/TriA-BioRE.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; **Document analysis**;
• **Computing methodologies** → **Natural language processing**.

KEYWORDS

Biomedical relation extraction, Biomedical literature, Domain-specific language models, Pair-aware representation, Triangular attention

*Corresponding authors.

BCB '23, September 3–6, 2023, Houston, TX, USA

ACM ISBN 979-8-4007-0126-9/23/09.
https://doi.org/10.1145/3584371.3612994

ACM Reference Format:
Lei Chen, Junhao Su, Tak-Wah Lam, and Ruibang Luo. 2023. Exploring Pair-Aware Triangular Attention for Biomedical Relation Extraction. In *14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '23), September 3–6, 2023, Houston, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3584371.3612994

1 INTRODUCTION

Biomedical relational facts are explicitly or implicitly hidden in a vast amount of biomedical literature, which have great significance in assisting clinical diagnosis, treatment, medical discovery, etc. However, extracting these valuable knowledge manually through the efforts of experts or researchers is becoming increasingly impractical, especially given the exponential growth of biomedical literature. To tackle this challenge, biomedical relation extraction (BioRE) has received growing attention in recent years from both academia and industry as a means of automatically extracting these relational facts from the unstructured biomedical literature. BioRE aims to identify the true relations between different biomedical entities [13], and many representative benchmark datasets have been built to facilitate the task, such as CDR [7] and GDA [12], which are annotated to predict the binary associations between Chemical and Disease concepts, Gene and Disease concepts, respectively, and BioRED [8], acting as a much more challenging dataset for predicting the multiple associations between Gene, Chemical, Disease and Variant concepts.

With the rapid advancement of deep learning techniques and the development of large-scale pretrained language models like BERT [1], the research of various natural language processing (NLP) tasks in general domain has achieved remarkable success and moved up to a new level. Simultaneously, the advent of domain-specific pretrained language models like BioBERT [6] and PubMedBERT [2] specifically customized for the biomedical domain has revolutionized BioRE task and achieved impressive performance gains by fully learning contextualized entity representations and effectively mining relational knowledge. Furthermore, other efforts like redesigning model structures or loss functions [10, 14] have further improved performance in BioRE. For example, ATLOP [14] attained better contextualized entity representations and training objectives by proposing a localized context pooling strategy and an adaptive

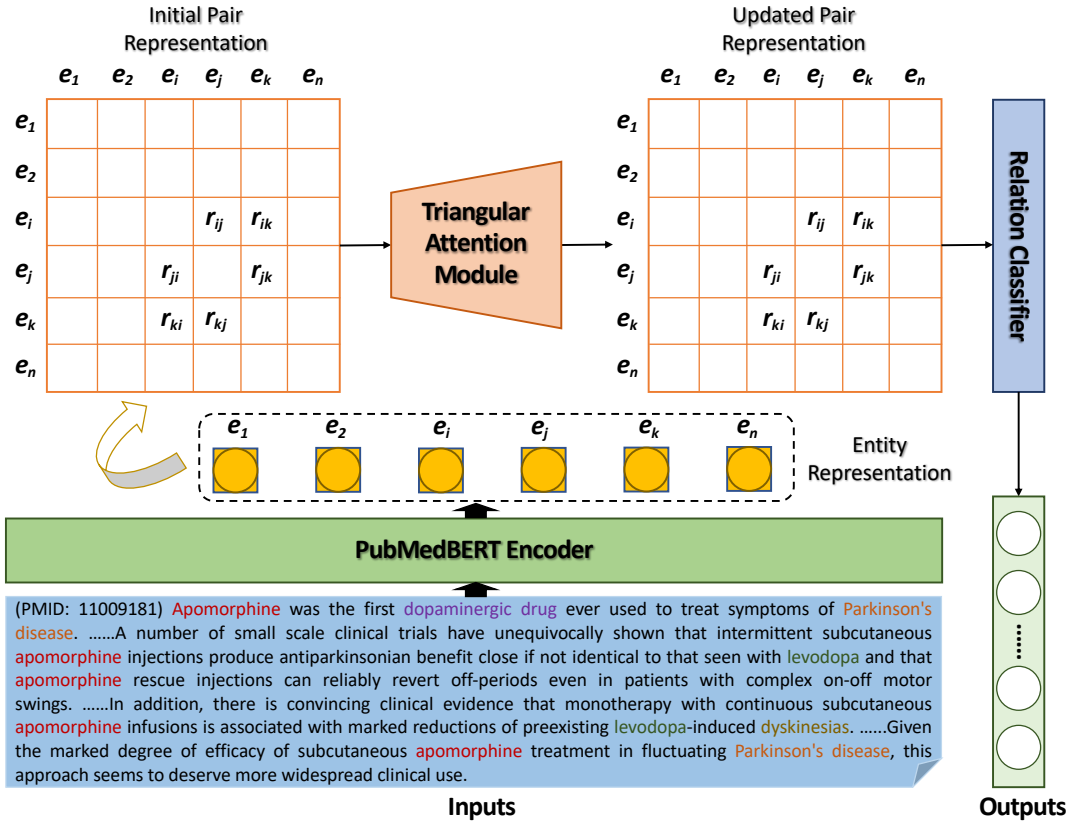

Figure 1: The architecture of TriA-BioRE framework. TriA-BioRE is mainly composed of three fundamental components: a PubMedBERT encoder, a triangular attention module, and a relation classifier.

thresholding loss. In [10], an enhanced adaptive focal loss was proposed as a replacement for the adaptive thresholding loss to address the issue of class imbalance between positive and negative samples.

Despite the remarkable progress achieved by previous studies in BioRE, there remains a critical challenge in learning high-quality entity pair representations for this task. Existing methods like BioBERT and PubMedBERT focusing on entity-level modeling have achieved improved contextualized entity representations. However, we posit that neglecting the high-level pair-aware representations learning is insufficient and can only achieve sub-optimal performance, particularly in the complicated context of the biomedical literature. Additionally, the importance of in-depth interaction of pair-level representations has been verified in many other tasks, such as protein representation learning and structure prediction [5, 9].

To tackle the aforementioned issue, we propose a novel **Tri**angular **A**ttention framework for **Bio**medical **R**elation **E**xtraction (called TriA-BioRE) to comprehensively exploring pair-aware representations in the biomedical domain. Specifically, we present a triangular attention module to enhance the pair-level modeling omnidirectionally for better BioRE performance. Concretely, after obtaining the contextualized entity representations through a powerful encoder like PubMedBERT, the initial pair representations are then further modeled by the triangular attention module, which encompasses two triangular multiplications utilizing outgoing and incoming edges, and two triangular self-attention operations centered on the starting and ending nodes, respectively. These triangular operations work in tandem to effectively capture the interdependency between different pairs and update the pair representations for accomplishing the BioRE task.

We conduct extensive experiments on three benchmark biomedical datasets (i.e., CDR [7], GDA [12] and BioRED [8]). Experimental results demonstrate that TriA-BioRE substantially outperforms strong competitors for BioRE.

2 OUR METHODOLOGY

Problem Definition. The goal of BioRE is to identify the correct relations between different biomedical entities. Formally, given a biomedical document D which consists of a set of biomedical entities $\{e_i\}_{i=1}^{n}$, BioRE aims to predict the true relations from $\mathcal{R} \cup \{NR\}$ between head and tail entity pairs $(e_h, e_t)_{h,t \in \{1...n\}, h \neq t}$, where $\mathcal{R}$ is a pre-defined set of relation types and NR represents for *No Relation*, e_h and e_t refer to head and tail entity respectively, and n is the number of entities. Note that an entity e_i may occur multiple times in D by entity mentions $\{m_j^i\}_{j=1}^{N_{e_i}}$, where N_{e_i} represents the number of entity mentions, and a relation exists between a head and tail entity pair (e_h, e_t) if it is expressed by any entity pair of

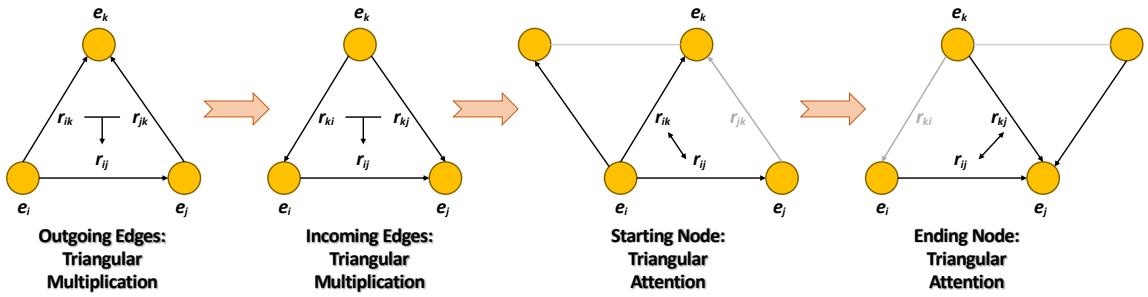

Figure 2: The workflow of Triangular Attention Module. In this core module, two triangular multiplications using outgoing and incoming edges, and two triangular self-attention operations around starting and ending nodes, are adopted to omnidirectionally capture the interdependency between different entity pairs.

their mentions. During inference stage, the true relation labels of all head and tail entity pairs $(e_h, e_t)_{h,t\in\{1...n\},h\neq t}$ need to be predicted.

The Overall Architecture. In this paper, we propose TriA-BioRE, a novel pair-aware triangular attention framework for BioRE. Figure 1 illustrates an overview of the TriA-BioRE framework, which mainly consists of three fundamental components: a PubMedBERT encoder, a triangular attention module, and a relation classifier. Specifically, in TriA-BioRE, the contextualized entity representations are first obtained by a powerful PubMedBERT encoder with modeling the whole biomedical document. Then, the initial pair representations are further modeled by a triangular attention module, which encompasses two triangular multiplications utilizing outgoing and incoming edges, and two triangular self-attention operations centered on the starting and ending nodes, respectively, to capture the interdependency between different pairs and update the pair representations. Finally, a relation classifier implemented with a feedforward neural network (FFN) is adopted to predict the true relation labels of all entity pairs. Next, we will describe each component in TriA-BioRE in detail.

2.1 TriA-BioRE

PubMedBERT Encoder. We adopt one of the most powerful domain-specific pretrained language models customized for the biomedical domain, i.e., PubMedBERT, as our encoder given its superior performance in recent studies [2, 8]. Given a biomedical document D of length l, we have $D = [x_t]_{t=1}^{l}$, and we add a special token "*" at the start and end position of each entity mention to mark the entities, following previous studies [10, 14]. Then, we utilize the PubMedBERT encoder to obtain the contextualized representations $\boldsymbol{H}$ of document D:

$$\boldsymbol{H} = \text{PubMedBERT}([x_1, \ldots, x_l]) = [h_1, \ldots, h_l] \tag{1}$$

We then take the representations of the special token "*" at the start position of the entity mentions as its embeddings, denoted as $\boldsymbol{h}_{m_j^i}$. For each entity e_i with entity mentions $\{m_j^i\}_{j=1}^{N_{e_i}}$, its entity representation $\boldsymbol{h}_{e_i}$ is calculated by a smoother logsumexp pooling [4] compared to max pooling operation:

$$\boldsymbol{h}_{e_i} = \log \sum_{j=1}^{N_{e_i}} \exp\left(\boldsymbol{h}_{m_j^i}\right) \tag{2}$$

Additionally, it is noteworthy that we also adopt the contextual pooling method from ATLOP [14] to obtain the context-enhanced entity representations $\boldsymbol{e}_i$, which is verified to be useful in BioRE. Then, the final entity pair representation $\boldsymbol{r}_{ht}$ for each head and tail entities (i.e., $\boldsymbol{e}_h$ and $\boldsymbol{e}_t$) is obtained by a feature combination through group bilinear pooling, following [14], which splits the entity representations into z equal-sized groups (e.g., $\boldsymbol{e}_h = \left[\boldsymbol{e}_h^1; \ldots; \boldsymbol{e}_h^z\right]$) and applies bilinear pooling function within the groups:

$$\boldsymbol{r}_{ht} = \sum_{i=1}^{z} \boldsymbol{e}_h^{i\top} \boldsymbol{W}_r^i \boldsymbol{e}_t^i + b_r \tag{3}$$

where $\boldsymbol{W}_r^i$ for $i = 1 \ldots z$ are learnable parameters, b_r is a bias term.

In existing methods, the final entity pair representation is directly used for relation classification, ignoring the high-level pair-aware representations learning, which is insufficient and can only achieve sub-optimal performance, particularly in the complicated context of the biomedical literature. To address this issue, we propose a novel triangular attention module to enhance the pair-level modeling omnidirectionally for better logical reasoning for BioRE.

Triangular Attention Module. To explore the high-level pair-aware representations effectively, we propose a novel triangular attention module to enhance the pair-level modeling omnidirectionally for better BioRE performance. As illustrates in Figure 2, triangular attention module consists of two triangular multiplications utilizing outgoing and incoming edges, and two triangular self-attention operations centered on the starting and ending nodes, respectively, to capture the interdependency between different pairs and update the pair representations.

First, we use an $\mathbb{R}^{n\times n\times d}$ matrix $\boldsymbol{P}$ to represent all head and tail entity pairs, and the diagonal of the $n \times n$ index is neglected, where n represents the number of entities, d is the embedding dimension. Then, the pair representation matrix $\boldsymbol{P}$ is regarded as a directed graph, with each entity $\boldsymbol{e}_i$ as a node and each pair representation $\boldsymbol{r}_{ij}$ as a directed edge. Particularly, we construct a triangle with edges $\boldsymbol{r}_{ij}$, $\boldsymbol{r}_{ik}$, and $\boldsymbol{r}_{jk}$, involving three different nodes $\boldsymbol{e}_i$, $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$, to update the pair representations. The first part in updating is two triangular multiplications utilizing outgoing and incoming edges, which are two symmetric operations. Specifically, a gating mechanism is firstly designed to dynamically choose information

Table 1: Dataset statistics (after preprocessing). Note that "# ET" and "# RT" are short for number of entity types and relation types respectively. And "# D", "Avg.# E" and "Avg.# R' are short for total number of documents, average number of entities and relations per document respectively.

Dataset	# ET	# RT	Train			Dev			Test		
			# D	Avg.# E	Avg.# R	# D	Avg.# E	Avg. # R	# D	Avg.# E	Avg.# R
CDR	2	2	500	6.9	10.9	500	6.7	10.5	500	6.8	10.8
GDA	2	2	23,353	4.9	5.7	5,839	4.8	5.7	1,000	4.7	5.2
BioRED	4	9	395	10.0	49.6	98	11.3	63.1	100	11.3	65.8

to be used to update the pair representation $\boldsymbol{r}_{ij}$:

$$\boldsymbol{g}_{ij} = \sigma\left(\boldsymbol{W}_g \boldsymbol{r}_{ij}\right) \tag{4}$$

where σ is the *sigmoid* function, $\boldsymbol{W}_g$ is a learnable weight matrix.

Then, the triangular multiplications utilizing outgoing edges (i.e., $\boldsymbol{r}_{ik} \cdot \boldsymbol{r}_{jk}$) and incoming edges (i.e., $\boldsymbol{r}_{ki} \cdot \boldsymbol{r}_{kj}$) are sequentially performed along with the gating filtering to accomplish the updates:

$$\begin{aligned} \widetilde{\boldsymbol{r}}_{ij}^{out} &= \boldsymbol{g}_{ij}^{out} \cdot \left(\boldsymbol{W}_{out}\left(\sum\nolimits_k \boldsymbol{r}_{ik} \cdot \boldsymbol{r}_{jk}\right)\right) \\ \widetilde{\boldsymbol{r}}_{ij}^{in} &= \boldsymbol{g}_{ij}^{in} \cdot \left(\boldsymbol{W}_{in}\left(\sum\nolimits_k \boldsymbol{r}_{ki} \cdot \boldsymbol{r}_{kj}\right)\right) \end{aligned} \tag{5}$$

where $\boldsymbol{W}_{out}$ and $\boldsymbol{W}_{in}$ are learnable parameters.

After the two triangular multiplications in updating, two another triangular self-attention operations centered on the starting and ending nodes are adopted to further capture the interdependency between the pair representations. Specifically, queries ($\boldsymbol{q}_{ij}$), keys ($\boldsymbol{k}_{ij}$) and values ($\boldsymbol{v}_{ij}$) are all derived by a linear projection from the corresponding pair representation $\boldsymbol{r}_{ij}$. And the self-attention weight $\boldsymbol{a}_{ijk}$ is calculated by all edges (i.e., $\boldsymbol{r}_{ik}$) sharing the same starting node $\boldsymbol{e}_i$, as well as modulated by the third edge information $\boldsymbol{t}_{jk}$ which is derived by a linear projection from the third edge $\boldsymbol{r}_{jk}$:

$$\boldsymbol{a}_{ijk}^{start} = \text{softmax}\left(\frac{\boldsymbol{q}_{ij}\boldsymbol{k}_{ik}{}^{\top}}{\sqrt{c}} + \boldsymbol{t}_{jk}\right) \tag{6}$$

where c is the channel dimension, and $\sqrt{c}$ is the scaling factor to avoid the large values of the inner product [11].

Then the updated pair representation $\widetilde{\boldsymbol{r}}_{ij}$ is obtained by the multiplications with the self-attention weights $\boldsymbol{a}_{ijk}$ and the values $\boldsymbol{v}_{ik}$, along with a gating filtering operation as well:

$$\widetilde{\boldsymbol{r}}_{ij}^{start} = \boldsymbol{g}_{ij}^{start} \cdot \sum\nolimits_k \boldsymbol{a}_{ijk}^{start} \boldsymbol{v}_{ik} \tag{7}$$

Similarly, triangular self-attention operation centered on the ending node is a symmetric operation of the above. And the corresponding self-attention weight $\boldsymbol{a}_{ijk}$ and updated pair representation $\widetilde{\boldsymbol{r}}_{ij}$ are formulated as:

$$\begin{aligned} \boldsymbol{a}_{ijk}^{end} &= \text{softmax}\left(\frac{\boldsymbol{q}_{ij}\boldsymbol{k}_{kj}{}^{\top}}{\sqrt{c}} + \boldsymbol{t}_{ki}\right) \\ \widetilde{\boldsymbol{r}}_{ij}^{end} &= \boldsymbol{g}_{ij}^{end} \cdot \sum\nolimits_k \boldsymbol{a}_{ijk}^{end} \boldsymbol{v}_{kj} \end{aligned} \tag{8}$$

Note that we add a residual connection after every update for robust performance [3]. Through such omnidirectional updates, we can obtain high-quality entity pair representations for better BioRE performance.

Relation Classifier. Finally, a relation classifier implemented with a feedforward neural network (FFN) is adopted to predict the true relation labels r of all entity pairs (e_h, e_t) based on the updated pair representations:

$$P\left(r \mid (e_h, e_t)\right) = \text{softmax}\left(\mathbf{W}_o \widetilde{\boldsymbol{r}}_{ht} + b_o\right) \tag{9}$$

where $\boldsymbol{W}_o$ is a learnable weight matrix, b_r is a bias term, and $\widetilde{\boldsymbol{r}}_{ht}$ is the final updated entity pair representation for each head and tail entities (i.e., $\boldsymbol{e}_h$ and $\boldsymbol{e}_t$).

Particularly, we adopt the adaptive focal loss [10] to optimize the whole TriA-BioRE framework, given its superior performance in tackling the class imbalance problem.

3 EXPERIMENTS

3.1 Experimental Datasets

We conduct extensive experiments on three benchmark biomedical datasets: CDR [7], GDA [12] and BioRED [8]. Specifically, CDR and GDA are annotated to predict the binary associations between Chemical and Disease concepts, Gene and Disease concepts, respectively, while BioRED is a much more challenging dataset for predicting the multiple associations between Gene, Chemical, Disease and Variant concepts. The detailed dataset statistics are reported in Table 1.

3.2 Baselines and Evaluation Metrics

To evaluate the effectiveness of TriA-BioRE, we compare it with several state-of-the-art baselines, including PubMedBERT [2], ATLOP [14], ATLOP (AFL) (i.e., enhanced ATLOP by replacing the original adaptive thresholding loss with a adaptive focal loss [10]). In addition, we adopt three widely-utilized classification evaluation metrics to measure the BioRE performance, including P (Precision), R (Recall) and F1 (F1 score) [12].

3.3 Quantitative Results

Table 2 reports the overall performance of TriA-BioRE and baselines on the three biomedical datasets. We can observe that TriA-BioRE consistently surpasses the compared models on all three datasets in terms of F1 score. For example, on CDR and GDA, TriA-BioRE obtains 0.44 and 0.77 improvements of F1 score respectively over the best baseline ATLOP (AFL). Particularly, on the much more challenging BioRED dataset, the improvement of F1 score is 1.23, which indicates the superiority of the pair-aware triangular attention learning compared to solely modeling entity-level representations.

Table 2: Overall results on the three datasets (i.e., CDR, GDA and BioRED) in terms of P (Precision), R (Recall) and F1 (F1 score).

Model	CDR			GDA			BioRED		
	P	R	F1	P	R	F1	P	R	F1
PubMedBERT	61.46	66.14	63.71	78.61	85.89	82.09	44.05	42.97	43.50
ATLOP	60.40	67.82	63.90	84.24	81.49	82.84	49.49	45.82	47.58
ATLOP (AFL)	61.14	68.48	64.60	83.01	82.96	82.98	55.24	44.11	49.05
TriA-BioRE	62.64	67.64	**65.04**	82.45	85.09	**83.75**	61.77	42.40	**50.28**
w/o Outgoing & Incoming	61.36	68.39	64.68	80.92	85.55	83.17	60.11	42.40	49.72
w/o Starting & Ending	58.12	73.17	64.78	84.07	81.89	82.97	54.97	45.25	49.63

3.4 Ablation Study

To assess the contribution of the proposed triangular attention module to the superiority of TriA-BioRE, we conduct an ablation study in terms of discarding the triangular multiplications utilizing outgoing and incoming edges (referred to as w/o Outgoing & Incoming), and the triangular self-attention operations centered on the starting and ending nodes (referred to as w/o Starting & Ending), respectively, as shown in the last two rows in Table 2. Note that our proposed TriA-BioRE will degenerate to the similar architecture of ATLOP (AFL) when removing the whole triangular attention module, resulting in a noticeable decrease of overall performance. Furthermore, when removing either the triangular multiplications utilizing outgoing and incoming edges or the triangular self-attention operations centered on the starting and ending nodes, the overall performance on all three datasets decreases significantly, demonstrating the importance and necessity to enhance the pair-level modeling omnidirectionally.

4 CONCLUSION

In this paper, we propose TriA-BioRE, a novel pair-aware triangular attention framework for biomedical relation extraction. Specifically, we present a triangular attention module with two triangular multiplications utilizing outgoing and incoming edges, and two triangular self-attention operations centered on the starting and ending nodes, respectively, together to enhance the pair-level modeling omnidirectionally. Extensive experiments on three biomedical datasets demonstrate that TriA-BioRE achieves substantially better results than the strong competitors for BioRE.

ACKNOWLEDGMENTS

R.L. was supported by Hong Kong Research Grants Council grants GRF (17113721) and TRS (T21-705/20-N), the Shenzhen Municipal Government General Program (JCYJ20210324134405015), the URC fund at HKU.

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[2] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)* 3, 1 (2021), 1–23.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[4] Robin Jia, Cliff Wong, and Hoifung Poon. 2019. Document-Level N-ary Relation Extraction with Multiscale Representation Learning. *arXiv preprint arXiv:1904.02347* (2019).

[5] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 7873 (2021), 583–589.

[6] 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 (2020), 1234–1240.

[7] Jiao Li, Yueping Sun, Robin J Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Thomas C Wiegers, and Zhiyong Lu. 2016. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database* 2016 (2016).

[8] Ling Luo, Po-Ting Lai, Chih-Hsuan Wei, Cecilia N Arighi, and Zhiyong Lu. 2022. BioRED: a rich biomedical relation extraction dataset. *Briefings in Bioinformatics* 23, 5 (2022), bbac282.

[9] Roshan M Rao, Jason Liu, Robert Verkuil, Joshua Meier, John Canny, Pieter Abbeel, Tom Sercu, and Alexander Rives. 2021. MSA transformer. In *International Conference on Machine Learning*. PMLR, 8844–8856.

[10] Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. 2022. Document-level relation extraction with adaptive focal loss and knowledge distillation. *arXiv preprint arXiv:2203.10900* (2022).

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[12] Ye Wu, Ruibang Luo, Henry CM Leung, Hing-Fung Ting, and Tak-Wah Lam. 2019. Renet: A deep learning approach for extracting gene-disease associations from literature. In *Research in Computational Molecular Biology: 23rd Annual International Conference, RECOMB 2019, Washington, DC, USA, May 5-8, 2019, Proceedings 23*. Springer, 272–284.

[13] Deyu Zhou, Dayou Zhong, and Yulan He. 2014. Biomedical relation extraction: from binary to complex. *Computational and mathematical methods in medicine* 2014 (2014).

[14] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-Level Relation Extraction with Adaptive Thresholding and Localized Context Pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*.