Learning to Leverage High-Order Medical Knowledge Graph for Joint Entity and Relation Extraction

Zhe Yang,Yi Huang*, and Junlan Feng*
JIUTIAN Team, China Mobile Research Institute
{yangzhe,huangyi,fengjunlan}@chinamobile.com

Abstract

Automatic medical entity and relation extraction is essential for daily electronic medical record (EMR) analysis, and has attracted a lot of academic attention. Tremendous progress has been made in recent years. However, medical terms are difficult to understand, and their relations are more complicated than general ones. Based on this situation, domain knowledge gives better background and contexts for medical terms. Despite the benefits of medical domain knowledge, the utilization way of it for joint entity and relation extraction is inadequate. To foster this line of research, in this work, we propose to leverage the medical knowledge graph for extracting entities and relations for Chinese Medical Texts in a collective way. Specifically, we propose to construct a high-order heterogeneous graph based on medical knowledge graph, which is linked to the entity mentions in the text. In this way, neighbors from the high-order heterogeneous graph can pass the message to each other for better global context representations. Our experiments on real Chinese Medical Texts show that our method is more effective than state-of-the-art methods.

1 Introduction

Medical text, e.g., electronic medical record (EMR), has been produced at a rapid speed and massive volume every day. Without any structured organization, this enormous volume of medical information is difficult to be read through by humans in a short time (Shang et al., 2021). Due to this situation, many researchers have recently paid great attention to the joint entity and relation extraction in the medical domain (Lai et al., 2021; Verlinden et al., 2021).

The challenge of joint entity and relation extraction in medical domain is that medical terms are usually difficult to understand due to the requirement of medical domain knowledge, especially for abbreviations of medical terms in the medical text. Even worse, relations between medical entities become even more complicated. Therefore, medical domain knowledge that could provide meaningful contexts and backgrounds is essential for the better extraction of medical entities and relations. Despite the advantages of medical domain knowledge, most previous works fail to use medical domain knowledge (Li et al., 2017; Xue et al., 2019; Pang et al., 2021; Luo et al., 2020). They solely rely on the local information in the medical text to extract entities and relations with language model (LM), which is insufficient for incomprehensible medical terms and complicated relations between entities.

Some recent works utilize medical knowledge for joint entity and relation extraction (Lai et al., 2021; Verlinden et al., 2021). However, both Lai et al. (2021) and Verlinden et al. (2021) simply align entity representation (node representation) from knowledge graph to local texts and fail to explicitly introduce the complicated relation contexts (edge representation) in the medical knowledge graph to enhance the deep representations of their involved entities. Huang et al. (2020) propose graph edge-conditioned attention network (GEANet) which integrates initial static relation embedding into attention mechanism for entity representation enhancement in medical knowledge graph. Nonetheless, it leaves out relational update during knowledge graph training process. Battaglia et al. (2018) propound graph network (GN) framework to update node and edge features iteratively within a heterogeneous graph. However, the edge representation update is based on the sender and receiver nodes information it links, which will bring about fluctuation since the amount of nodes is far outweighed that of edge types.

Therefore, we propose a method to fix these issues by providing additional relation contexts from

*Corresponding authors

Findings of the Association for Computational Linguistics: ACL 2023, pages 9023–9035
July 9-14, 2023 ©2023 Association for Computational Linguistics

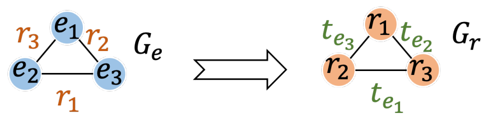

Figure 1: An example of the high-order heterogeneous graph: the left is original first-order graph searched directly from knowledge graph with e_i and r_i as entity and relation, respectively, and the right is its converted second-order graph where t_{e_i} means the type of entity e_i.

medical knowledge graph to enrich the deep representations of entity mentions. Specifically, we propose to construct a high-order heterogeneous graph, e.g., Fig. 1, to provide meaningful global contexts for its linked entity mentions. In Fig. 1, we denote G_e as the standard first-order graph with entities as nodes and relations as edges, and represent G_r as G_e's converted second-order graph with relations as nodes and entity types as edges. For every relation pair of an entity in G_e, e.g., r_3 and r_2, we link an edge, i.e., the entity type t_{e_1}, in G_r connecting them. In this way, both message passing of entities via different relations and message propagation of relations via different entity types can be well diffused in the global graph structure. After extracting the high-order heterogeneous graph from the medical knowledge graph, we fuse the entity and relation representations in the global context obtained from the high-order heterogeneous graph with the local information extracted from the medical text.

To summarize, our contributions are:

• We propose a high-order graph modeling method for knowledge fusion, which treats text related sub-graph as the first-order graph with entities as nodes and its converted graph as the second-order one with relations as nodes. We update the hidden representations of nodes in the two order graphs separately as the entity/relation representations for knowledge graph.

• We present a knowledge-enhancement method for medical text encoding, which boosts the entity representation of the first-order graph with the feedback of the second-order relation representation. And it further enhances the encoding of the entity mentions from medical text for joint extraction.

• We have performed substantial experiments against existing methods. Our evaluation results on real medical datasets verify that our method is more effective than state-of-the-art methods.

The rest of this paper is organized as follows. Section 2 discusses related work. In Section 3, we introduce the typical algorithms for text and graph representation and present the proposed method for knowledge-enhanced joint extraction. Section 4 shows the evaluation results on two datasets with compared to some other advanced methods. We conclude in Section 5.

2 Related Work

There are two categories of entity and relation extraction methods: pipeline-based methods and joint extraction methods.

Pipeline-based entity and relation extraction methods: These work usually first extract entities as outputs, then extract relations for the returned entities. The drawback of pipeline-based extraction methods is that the errors of entity extraction may be accumulated when extracting relations for the already returned entities. For example, Zhong and Chen (2021) put forward an extra encoder and fuse the entity type information to enhance entity pair representation during the relation extraction task.

Jointly entity and relation extraction methods: Some recent work extracts entity and relation in a collective way to overcome the accumulated error problem in the pipeline-based methods. There are two kinds of standard extraction methods, which are non-knowledge-enhanced and knowledge-enhanced methods.

Most joint entity and relation extraction methods ignore the domain knowledge. Wang et al. (2018) utilize a novel graph scheme to solve the problem. Luan et al. (2018) apply multi-task method to optimize entities, relations, and coreference simultaneously.Bekoulis et al. (2018) further propose multi-context based adversarial training method. Luan et al. (2019) utilize dynamic span graphs to form a general framework. Fu et al. (2019) model text as relational graphs for joint extraction. Zhao et al. (2021) model dense cross-modal interactions for joint extraction. Wang and Lu (2020) apply table-sequence encoders to extract jointly. Lin et al. (2020) apply neural model for information extraction with global features. Recently, Eberts and Ulges (2020) propose a span-based method with transformer pre-training. Further, Ji et al. (2020) apply span-based method with attention-based span-specific and contextual semantic representations. Moreover, Wei et al. (2020) propose a cascade binary tagging framework for entity and relation extraction. Yan et al. (2021) propose a partition filter network for joint entity and relation extraction. Despite that these works extract entity and relation in a joint way, they are not designed for medical text. While some previous works are designed for medical text, they ignore the help of medical domain knowledge when modeling (e.g., Li et al., 2017; Xue et al., 2019; Pang et al., 2021; Luo et al., 2020). Only a few works incorporate medical domain knowledge to enhance the contexts of the medical text for joint entity and relation extraction (Lai et al., 2021; Verlinden et al., 2021). Among them, Lai et al., 2021 extract entity and relation jointly with knowledge-enhanced collective inference. Verlinden et al. (2021) inject knowledge base information into entity and relation extraction and coreference resolution simultaneously. However, both Lai et al. (2021) and Verlinden et al. (2021) fail to explicitly incorporate complicated relation contexts for their involved entities. Thus the interaction between entity and relation extraction will not be captured. Our method overcomes this drawback by providing relation contexts when modeling deep representations of entities.

Most of the current work does not use domain knowledge, or only uses the entity context in domain knowledge, and does not include the relation context. While we explicitly model the entity context and relation context as the important context information for entity span, which improves the results of joint extraction.

3 Proposed Method

Fig. 2 shows the framework of our method. It first extracts the high-order heterogeneous graph from the medical knowledge graph (Section 3.1), and learns entity and relation representation from the global context (Section 3.2). After that, we learn the representation of entity mentions from the local context in the medical text (Section 3.3). The fusion of entity representations from both global and local contexts and its corresponding relation representation (Section 3.4) is utilized to extract entities and relations in a collective way (Section 3.5).

3.1 High-order Heterogeneous Graph Extraction

The challenge is that there are multiple relations in the heterogeneous graph, and modeling entities together with their contextual relation information is non-trivial. We propose to construct a high-order heterogeneous graph from the knowledge graph for the medical text, such that the text representation contains additional global knowledge contexts of related entities and relations.

The high-order heterogeneous graph comprises of first-order graph (a text related sub-graph of original knowledge graph) and its converted second-order graph. Given a piece of medical text expressed as words $t = (w_0, w_1, ..., w_i, ..., w_k)$ and a domain knowledge graph $G = (V, \varepsilon, E)$, the goal of high-order heterogeneous graph extraction is to obtain $G_e = (V_e, E_e)$ and $G_r = (V_r, E_r)$. Specifically,

$$\begin{aligned} V_e &= V \cap \{w_i\}, \\ (e_i, r_{e_i,e_j}, e_j) &\in \varepsilon, \\ r_{e_i,e_j} &= E_e(e_i, e_j), \end{aligned} \tag{1}$$

where $v_{e_i}, v_{e_j} \in V_e$, $r_{e_i,e_j} \in E_e$ and ε is the set of all triplets $\{(e_h, r_{e_h,e_t}, e_t)\}$ in knowledge graph. And

$$\begin{gathered} V_r \subseteq E_e, \\ E_r \subseteq T(V_e), \\ \mathbb{I}(E_r(E_e(e_i, e_j), E_e(e_l, e_m))) = \mathbb{I}(e_i = e_m \vee e_j = e_l), \end{gathered} \tag{2}$$

where $T(.)$ means the type of an entity. $\mathbb{I}$ is the indicator function and $\mathbb{I}(E_r(.))$ measures whether the edge $E_r(.)$ in second-order graph exists.

Hence, it first searches through the medical knowledge graph to extract all triplets as the first-order graph pertaining to current medical text, and converts it to the second-order graph with nodes and edges switched. To be detailed, we traverse nodes, i.e., entities, in the first-order graph and collect their **in** and **out edges**, i.e., relations. Concretely, when it is the head entity in a triplet, we call the corresponding relation the **out relation**, conversely, the **in relation**. For any in-out relation pair of an entity, we consider relations as nodes for the second-order graph, and the type of the corresponding entity as the edge linking them. As it can be seen in the *High-Order Heterogeneous Graph Extraction Module* part of Fig. 2, a Chinese entity “鹅口疮(thrush) ” is revolved around by relations of “ 可能疾病(possible disease) ”, “并发症(complication) ” and “传染方式(mode of infection) ” in the first-order graph, during which the first two are in relations, and the other is the out relation. For any in-out relation pair, e.g., “并发症-传染方式”, we take them as two nodes in the

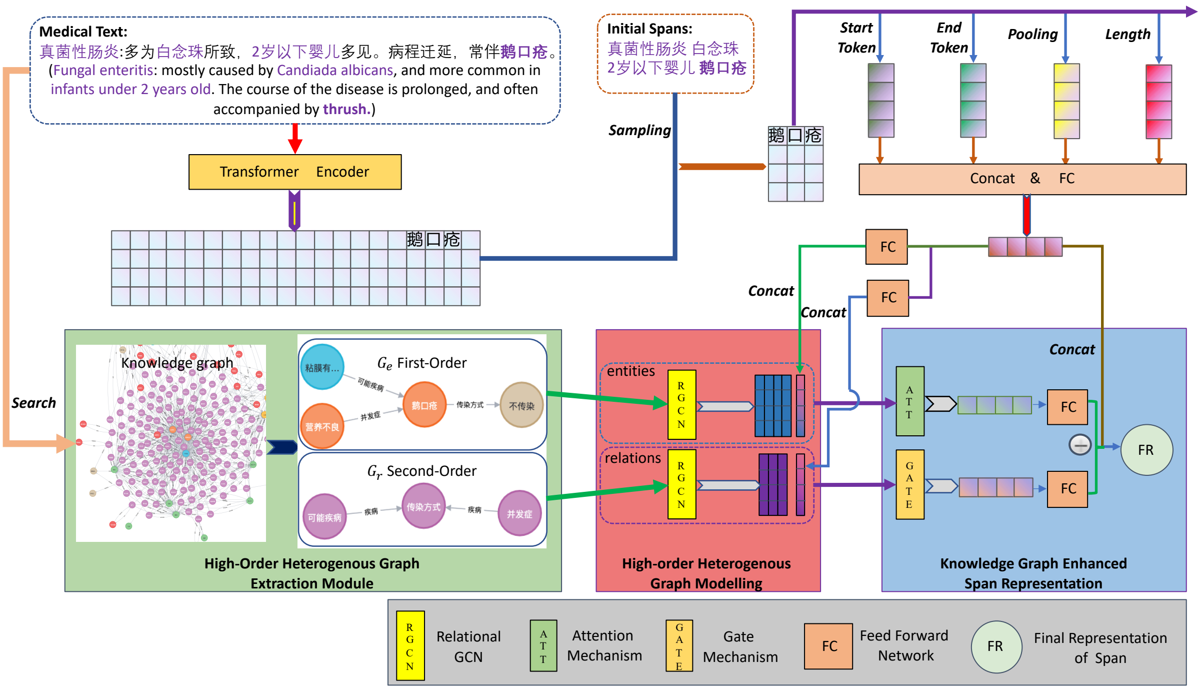

Figure 2: The overall framework of our model. Details of each step are from Section 3.1 to 3.5.

converted second-order graph and the type of the entity "鹅口疮", i.e., "疾病(disease)" as the linking edge. As a counter example, "可能疾病" and "并发症" are all **in relations** for entity "鹅口疮", therefore, no edge exists to link them. We merely link the in-out relation pair on account of the message flowing direction, i.e., from head entity to tail entity.

This high-order heterogeneous graph we propose is distinguished from the node layer and edge layer in Jiang et al. (2020), which lies in two aspects:

(1) We only consider the in-out relation pairs as edges for second-order graph, which better reflects the direction of information flow, while Jiang et al. (2020) links all relation pairs.

(2) The edge in second-order graph represents the entity type in our method, however, it remains as the entity by authors of Jiang et al. (2020), where the number of entities is very huge and will not be feasible for GNN to learn.

3.2 High-Order Heterogeneous Graph Modeling

Our idea is to propagate messages among both the first-order graph G_e and the converted second-order graph G_r in Fig. 2, in order to capture the complex global information from knowledge graph. Hence, entity mentions in the medical text can integrate their local information with the related global contexts for the better encoding.

We first propagate message among the standard first-order graph G_e with relational graph convolutional network (RGCN) (Schlichtkrull et al., 2018). We apply TransE (Bordes et al., 2013) as the initialization of the embedding for each node v_i. The embedding of each node v_i, i.e., entity, can be updated as:

$$v_i^{l+1} = \mathrm{ReLU}(\mathrm{U}^l v_i^l + \sum_{k \in R} \sum_{v_j \in N_i^k} (\frac{1}{|N_i^k|} \mathrm{U}_k^l v_j^l)) \,, \tag{3}$$

where v_i^l is the embedding of the node v_i at layer l. N_i^k is the set of neighbors of v_i under relation k. U^l is the trainable parameter at layer l. U_k^l is relation specific weighted parameter.

We further model the global contexts in the converted second-order graph G_r. Different from the first-order graph G_e, the nodes and edges in second-order graph represent relations and entity types respectively. We pass messages among the neighbors of the node, i.e., relation r_i, via different entity types in graph G_r. Then the information among multiple entity types of relation r_i can be summarized. Here the initialization of the embedding for each relation r_i is also obtained by the TransE model. The embedding of each node, i.e., relation r_i, is integrated by the deep representations of its neighbors as:

$$r_i^{l+1} = \text{ReLU}(\text{O}^l r_i^l + \sum_{t \in T} \sum_{r_j \in N_i^t} (\frac{1}{|N_i^t|} \text{O}_t^l r_j^l)) , \quad (4)$$

where r_i^l is the embedding of the relation r_i at layer l. N_i^t is the set of neighbors of relation r_i under the entity type $t \in T$, while $|N_i^t|$ is the number of neighbors of r_i. O^l is the trainable parameter at layer l. And ReLU is the activation function.

The high-order heterogeneous graph modeling provides deep representations for the entities and relations. Therefore, complicated relation information can be well preserved for the involved entities in the modeling of Section 3.5.

3.3 Spans Representation with Transformer Encoder

After modeling the high-order heterogeneous graph extracted from the medical knowledge graph, we then model the local contexts in the medical text. The medical text is organized here as tokens $t = (x_0, x_1, ..., x_i, ..., x_{n-1})$, and a successive sequence $(x_{s_i}, ..., x_{e_i})$ in text is a span which means an entity mention for medical text. Each span s_i is modeled as:

$$s_i = g_s([x_{s_i}, x_{e_i}, \hat{x}_i, \phi(s_i)]) , \quad (5)$$

where x_{s_i} denotes the token level embedding from the transformer encoder, e.g., BERT, of the start of span s_i, while x_{e_i} represents the token level embedding of the end of span s_i. $\hat{x}_i$ is an attention-weighted sum of the token representations in the span. And $\phi(s_i)$ is a feature vector modeling the length of s_i. g_s is a feed-forward neural network (Lee et al., 2017).

Then a span-based GCN is applied on the graph with spans as nodes and relations between spans as edges:

$$h_i^{l+1} = h_i^l + f_{span}^l(\text{ReLU}(f_s^l(h_i^l), f_{s'}^l(h_i^l))) , \quad (6)$$

and

$$f_s^l(h_i^l) = \sum_{s_j \in s_{set}, j \neq i} \sum_{k \in R} r_{ij}[k](W_k h_j^l + b_k),$$

$$f_{s'}^l(h_i^l) = \sum_{s_j \in s_{set}, j \neq i} \sum_{k \in R} r_{ji}[k](W'_k h_j^l + b'_k), \quad (7)$$

$$r_{ij}[k] = Softmax(f_r([s_i, s_j, s_i \circ s_j]))[k],$$

where h_i^{l+1} is the deep representation of s_i at the layer $l+1$, $f_{span}$and f_r are feedforward neural networks. f_s^l and $f_{s'}^l$ are bidirectional GCN (Marcheggiani and Titov, 2017; Fu et al., 2019) for h_i^l, $r_{ij}[k]$ measures the relation score for relation k between spans of s_i and s_j, $\circ$ denotes the element-wise multiplication. And h_i^0 is initialized as s_i.

3.4 Knowledge Graph Enhanced Span Representation

After applying span-based GCN, we obtain the hidden representation h_i of span s_i derived from local context, e.g., the medical text. Then we apply an attention mechanism (Lai et al., 2021) to integrate the deep representation of hidden representation h_i of span s_i from the local medical text with the deep representation of entities v_i from the global contexts in the first-order graph as f_{i_e}:

$$f_{i_e} = W_{i_e} f_{c_e}(h_i) + \sum_{v_j \in C(s_{i_e})} W_{ij_e} f_v(v_j) , \quad (8)$$

where $C(s_{i_e})$ is the candidate set of entities corresponding to span s_i in the dual heterogeneous graph. And $f_{c_e}(h_i)$ and $f_v(v_j)$ are the transformed representation of h_i and v_j by two feedforward neural networks. And W_{i_e} and W_{ij_e} are attention scores of the two transformed representations:

$$W_{i_e} = \frac{\exp(\alpha_{i_e})}{(\exp(\alpha_{i_e}) + \sum_{v_j \in C(s_i)} \exp(\alpha_{ij_e}))} , \quad (9)$$

and

$$W_{ij_e} = \frac{\exp(\alpha_{ij_e})}{(\exp(\alpha_{i_e}) + \sum_{v_j \in C(s_i)} \exp(\alpha_{ij_e}))} . \quad (10)$$

Here α_{ij_e} and α_{i_e} are importance scores of the transformed entity representation v_j and the transformed span representation h_i to the span representation h_i:

$$\alpha_{ij_e} = f_{\alpha_e}([h_i, f_v(v_j)]) , \quad (11)$$

and

$$\alpha_{i_e} = f_{\alpha_e}([h_i, f_{c_e}(h_i)]) , \quad (12)$$

where f_{α_e} is a feedforward neural network.

Next, we fuse the deep representations of relations r_i from the global contexts in the second-order graph G_r with the deep representation h_i of span s_i from the local medical text. Distinguished from first-order graph fusion in equation 8, we argue that the corresponding relations in second-order graph may be irrelevant to the span depending on current medical text, and arouse disproportionate or even noisy aggregating representations for local context. Therefore, a selective gate mechanism (Li et al., 2020) is utilized to perform this fusion as f_{i_r}:

$$f_{i_r} = g_{i_r} f_{c_r}(h_i) + \sum_{r_j \in C(s_{i_r})} g_{ij_r} f_r(r_j), \quad (13)$$

where $C(s_{i_r})$ is the candidate set of relations in the second-order heterogeneous graph for the span s_i. To be detailed, $C(s_{i_r})$ is a subset of relations from knowledge graph triplets where s_i holds as head or tail entity. And $f_{c_r}(h_i)$ and $f_r(r_j)$ are the transformed representation of h_i and r_j by two feedforward neural networks. And g_{i_r} and g_{ij_r} are gate scores of the two transformed representations:

$$g_{i_r}=\sigma(W_2(\text{ReLU}(W_1[f_{c_r}(h_i), h_i]+ b_1)+b_2), \quad (14)$$

and

$$g_{ij_r}=\sigma(W_2(\text{ReLU}(W_1[f_r(r_j), h_i]+b_1)+b_2), \quad (15)$$

where σ is the *Sigmoid* activate function which maps the results into the interval of $(0, 1)$.

Then, the fused representation f_i can be modeled as:

$$f_i = (W_{sum_e} f_{i_e} + W_{sum_r} f_{i_r})||h_i \ , \quad (16)$$

which means an integration of deep representation among entities, relations and the span. W_{sum_e} and W_{sum_r} serve as feedforward neural networks. $||$ means concatenation operation.

3.5 Collective Entity and Relation Extraction

Finally, we map the integrated representation f_i to the entity type space as:

$$e_i = \text{Softmax}(g_e(f_i)) \ , \quad (17)$$

where g_e is a feed-forward neural network to map f_i to the entity space.

Similarly, the relation between span i and j can be mapped to the relation type space as:

$$r_{ij} = \text{Softmax}(g_r(f_i, f_j)) \ , \quad (18)$$

where g_r is a feed-forward neural network to map f_i and f_j to the relation space.

The training of collectively extracting entity and relation can be optimized by minimizing the loss function as:

$$\mathcal{L} = \mathcal{L}_e + \mathcal{L}_r \ , \quad (19)$$

where $\mathcal{L}_e$ represents the cross-entropy loss of entities, and $\mathcal{L}_r$ denotes the cross-entropy loss of relations.

Different from the pipeline way that optimizes entities and relations in separate steps, our method conducts the optimization in an end-to-end mode. In this way, the errors of extracting entities and relations can be reduced collectively.

4 Experiments

In this section, we evaluate our model with extensive experiments. We first show the experimental setup, which contains datasets, baselines for comparison, and evaluation metrics. Then we evaluate the effectiveness of our model and baselines.

4.1 Experimental Setup

4.1.1 Datasets

We evaluate our method on several real medical datasets with one medical knowledge graph dataset.

Chinese Medical Text datasets: We evaluate our model on three Chinese medical text datasets.

• The first medical text dataset is the **CHIP-2020** dataset [1], which contains 17,924 sentences from biomedical Chinese text that captures relations between medical entities (Guan et al., 2020). The dataset includes a pediatric labeled corpus for hundreds of common diseases. The training and testing splits contain 14,339 and 3,585 sentences, respectively.

• The next medical dataset is the **CHIP-2022** dataset [2], which contains 1000 samples and is split into 850/150 for training and testing set, respectively. CHIP2022 aims to extract casual, entailment and conditional relations between entities whose type are not distinguished. In the experiments, we merely trace out the previous two types of relations for joint extraction.

• The third medical text dataset is the **DiaKG** dataset [3], which contains sentences from diabetes text that reflects relations between medical entities (Chang et al., 2021). The dataset comes from 41 diabetes guidelines and consensuses from authoritative journals in China. The diabetes text contains 22,050 entities and 6,890 relationships. Finally, the prepossessed training and testing splits contain 1,170 and 239 sentences, respectively.

[1] http://cips-chip.org.cn/2020/eval2
[2] http://cips-chip.org.cn/2022/eval2
[3] https://tianchi.aliyun.com/dataset/dataDetail?dataId=88836

Model	CHIP-2020			CHIP-2022			DIAKG		
	Entity	Relation	Overall	Entity	Relation	Overall	Entity	Relation	Overall
PFN (Yan et al., 2021)	72.2	52.7	62.5	59.3	**42.0**	**50.7**	63.0	44.5	53.8
CASREL (Wei et al., 2020)	66.1	49.5	57.8	56.5	39.1	47.8	50.4	31.0	40.7
SPERT (Eberts and Ulges, 2020)	73.5	50.1	61.8	57.8	38.7	48.3	65.5	29.1	47.3
KB-graph (Verlinden et al., 2021)	73.9	61.8	67.9	65.0	32.5	48.8	69.3	51.1	60.2
KECI (Lai et al., 2021)	74.3	61.1	67.7	63.4	32.6	48.0	72.6	**55.1**	63.8
KECI_nnconv (Li et al., 2017)	74.2	61.7	68.0	63.5	32.4	48.0	73.2	54.1	63.7
KECI_gea (Huang et al., 2020)	75.6	61.0	68.3	61.8	32.5	47.2	73.1	**54.9**	64.0
KECI_gn (Battaglia et al., 2018)	74.7	60.5	67.6	64.0	32.4	48.2	73.9	54.5	64.2
Our Model	**76.5**	**61.8**	**69.2**	**66.7**	32.5	49.6	**74.2**	54.4	**64.3**

Table 1: Evaluation results (%) on CHIP-2020 & CHIP-2022 & DiaKG datasets.

Chinese Medical knowledge graph dataset: We extract Chinese medical triplets from the public knowledge graph dataset [4], which contains disease-related entities and disease-related triples. After preprocessing, we obtain 215,745 triplets for the medical knowledge graph. These triplets contain 16,735 entities and 13 relations, and the amount of entity types is 9.

4.1.2 Baselines for comparison

We compare our model with state-of-the-art joint extraction methods, variant versions of knowledge-enhanced baseline included.

• **PFN** (**P**artition **F**ilter **N**etwork) is an advanced joint extraction method that does not utilize a knowledge graph. It segments the encoder into entity extraction and relation extraction parts, and accomplishes NER-specific and relation-specific tasks with shared part separately (Yan et al., 2021).

• **CASREL** (**Cas**cade Binary Tagging Framework for **Rel**ational Triple Extraction) is an advanced joint extraction method without knowledge graph enhancement. It models triplets extraction into head entity extraction as well as relation and tail entity extraction with fused head entity representation. The two parts share the same encoder for a joint extraction task (Wei et al., 2020).

• **SPERT** (**Sp**an-based Joint **E**ntity and **R**elation Extraction with **T**ransformer Pre-training) is also an advanced joint extraction method without knowledge graph context. It attaches entity labels for spans derived from the text, and traversed span pairs for relation judgement (Eberts and Ulges, 2020).

• **KB-graph** is an advanced joint extraction method, which injects medical knowledge into entity and relation extraction and coreference resolution simultaneously (Verlinden et al., 2021).

• **KECI** (**K**nowledge-**E**nhanced **C**ollective **I**nference) is an advanced joint extraction method, which extracts entity and relation collectively with knowledge-enhanced collective inference. It utilizes RGCN algorithm for knowledge graph representation (Lai et al., 2021). Compared with the proposed method, KECI merely models the first-order graph for knowledge fusion.

• **KECI_nnconv** is the variant version of KECI (Lai et al., 2021) in which the NNConv (Gilmer et al., 2017) is utilized instead of RGCN. Different from RGCN, NNConv fuses the initial edge feature for message propagation and further enhances the node representation, which is similar to GEANet.

• **KECI_gea** is one variant version of KECI (Lai et al., 2021) that the RGCN part is replaced by GEANet (Huang et al., 2020) which is similar to NNCONV.

• **KECI_gn** is also one variant version of KECI (Lai et al., 2021) where the node and edge features are simultaneously derived from the graph net (GN) framework (Battaglia et al., 2018).

4.1.3 Evaluation metrics

We evaluate methods by Micro-F1 scores for entity and relation extraction. Also, we use the average Micro-F1 scores of entities and their relations in each medical text as the overall scores for each medical text.

4.1.4 Implementation details

We implement our method with PyTorch and MedBERT-kd Transformer [5] which is based on the structure of BERT and trained on a mount of Chinese clinic texts. The number of parameters of our model is 135,557,734. We conduct hyper-parameter tuning by using a Bayesian optimizer for all the methods on real datasets. The scopes of hyper-parameters are: {16, 32} for batch size, {2e-5, 3e-5, 4e-5, 5e-5} for lower learning rate, {1e-4, 2e-4, 5e-4} for upper learning rate.

A coarse-to-fine training process is applied during training process. Firstly, we leave out the

[4] http://www.openkg.cn/dataset/medical

[5] https://huggingface.co/trueto/medbert-kd-chinese

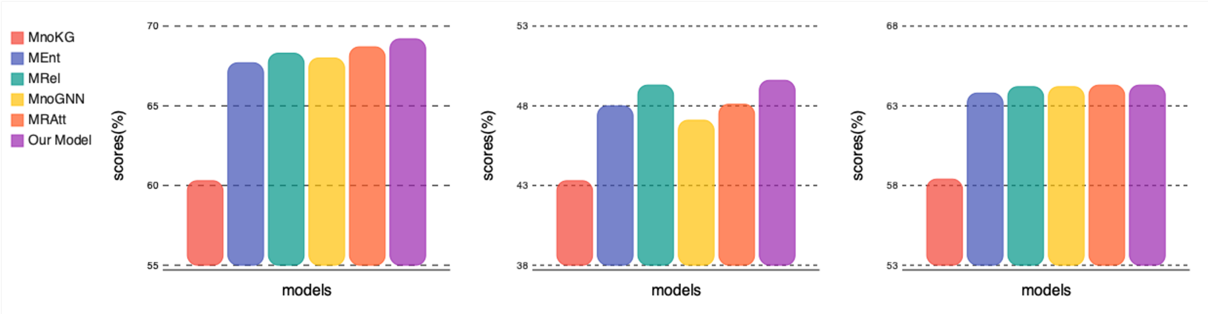

Figure 3: The ablation study results w.r.t. overall Micro-F1 scores on CHIP-2020/CHIP2022/DiaKG datasets.

knowledge graph fusion part and only employ hidden representations of spans from the local context to train a better downstream task-specific BERT structure. After that, global information from knowledge graph representation is added for a precise training.

We evaluate all models with GPUs on the JIUTIAN Platform of China Mobile Research Institute.

4.2 Evaluation results

We evaluate the effectiveness of all methods on three real datasets in Table 1, where we show the Micro-F1 values of both entities and relations. We also show the averaged overall Micro-F1 values. From Table 1, we make the following observations:

(1) KECI outperforms other joint extraction methods which do not involve medical knowledge graph. For example, KECI has 9.9%, 0.2% and 23.1% improvements compared with CASREL on CHIP-2020, CHIP-2022 and DiaKG datasets, respectively. It shows that the medical knowledge graph is essential for providing more contexts in joint entity and relation extraction.

(2) Our model works better than KECI in overall values. Specifically, our model achieves 1.5%, 1.6% and 0.5% improvements compared with the KECI model on CHIP-2020, CHIP-2022 and DiaKG datasets, respectively. It verifies that high-order graph context could provide more information for text modeling. Moreover, before adding relation context, as shown in Table 1, the Micro-F1 score of the entity for KECI is 74.3 % on CHIP-2020 dataset, while our method uses relation context and Micro-F1 is increased to 76.5 %. On average, there are 2.2%, 3.3%, and 1.6% improvements on the three datasets respectively after adding relation contexts. It can be seen that the relation context does have a significant effect on improving the entity value.

(3) Results in **CHIP-2022** show that our model has a little poorer performance than that of **PFN** in overall score, however, we still prevail much in **entity** result. This is attributed to the fact that the relation types in **CHIP-2022** dataset are all logical types, such as casual and entailment relations, which varies considerably from medical knowledge graph.

We further make detailed evaluation on all the entity types which can be seen in table 2 and 3. Generally, our model achieves a higher F1 score among entity types, especially in the less-train-instance targets, e.g., "部位(part) ", "其他治疗(other therapy) ", "其他(others) " in CHIP-2020 dataset and "Amount", "Method", "Pathogenesis" in DiaKG dataset, our model performs almost above 4% better than KECI (Lai et al., 2021). When mentioning CHIP-2022 data, we neglect the entity type results owing to its lack of discrimination between entity types.

We also make comparisons between proposed high-order heterogeneous graph modeling method and some typical graph neural networks (GNN). We can conclude that the high-order graph modeling for both entity and relation update is more efficient than a single order heterogeneous graph modeling for entity representation, even with the extra relation update. Details are as follows:

(1) The proposed model exceeds KECI_nnconv, KECI_gea and KECI_gn by 1.2%, 0.9%, 1.6%, respectively, in CHIP-2020 dataset.

(2) The proposed model surpasses KECI_nnconv, KECI_gea and KECI_gn by 1.6%, 2.4%, 1.4%, respectively, in CHIP-2022 dataset.

(3) The proposed model beats KECI_nnconv, KECI_gea and KECI_gn by 0.6%, 0.3%, 0.1%, respectively, in DiaKG dataset.

Entity Type	KECI(Lai et al., 2021)			Our Model		
	P	R	F1	P	R	F1
surgery	55.78	54.30	55.03	47.34	64.90	54.75
inspection	65.75	61.03	63.30	58.07	72.31	64.41
epidemiology	60.76	58.22	59.46	52.92	73.26	61.45
disease	70.04	84.52	76.60	79.95	80.24	80.09
symptom	71.70	63.12	67.14	68.00	72.74	70.29
sociology	66.83	46.64	54.94	60.14	57.79	58.94
medicine	61.27	75.68	67.72	66.84	75.78	71.03
part	72.93	46.41	56.72	60.35	66.27	63.17
prognosis	-	-	-	-	-	-
other-therapy	47.70	33.24	39.18	47.67	38.78	42.77
others	-	-	-	100.00	5.83	11.02

Table 2: Entity evaluation details (%) on CHIP-2020 dataset. (- means the result is zero.)

Entity Type	KECI(Lai et al., 2021)			Our Model		
	P	R	F1	P	R	F1
ADE	68.42	54.17	60.47	77.55	52.78	62.81
Amount	20.00	17.65	18.75	26.67	23.53	25.00
Anatomy	79.31	92.00	85.18	82.51	92.00	87.00
Class	87.13	60.69	71.55	86.27	60.69	71.25
Disease	78.22	77.49	77.85	73.40	82.60	77.73
Drug	65.02	75.24	69.76	66.80	77.62	71.80
Method	-	-	-	36.84	50.00	42.42
Pathogenesis	-	-	-	22.22	25.00	23.53
Reason	30.00	33.33	31.58	-	-	-
Symtom	48.15	40.62	44.07	52.17	37.50	43.63
Test	43.33	61.90	50.98	50.00	66.67	57.14
Test_items	59.54	60.94	60.23	53.99	68.75	60.48
Treatment	42.86	57.69	49.18	38.24	50.00	43.34
Operation	-	-	-	-	-	-

Table 3: Entity evaluation details (%) on DiaKG dataset. (- means the result is zero.)

4.3 Ablation Study

We conduct an ablation study on our model to evaluate the effectiveness of its main modules in Figures 3. Specifically, we compare our model with the following variants w.r.t. the overall score of each medical text, i.e., average Micro-F1 score of entities and relation in each medical text.

• **MnoKG** (**M**odel without **K**nowledge **G**raph) ignores knowledge graph for triplets extraction.

• **MEnt** (**M**odel with **Ent**ity) merely fuses first-order graph into medical text for joint extraction.

• **MRel** (**M**odel with **Rel**ation) merely fuses second-order graph into medical text for extraction.

• **MnoGNN** (**M**odel without **GNN** update) fuses initial node representations of high-order graph from TransE into text without GNN to update.

• **MRAtt** (**M**odel with **R**elation **Att**ention fusion) is one variant version of our model, which uses an attention mechanism in Equation 20 to replace Equation 16 for relation fusion:

$$f_{i_r} = W_{i_r} f_{c_r}(h_i) + \sum_{r_j \in C(s_{i_r})} W_{ij_r} f_r(r_j) \ , \quad (20)$$

where W_{i_r} and W_{ij_r} are similar to 9 and 10, respectively.

From the Figures, we make the following observations:

(1) Our model beats **MnoKG** on all real datasets. In particular, the performance improvements of our model are 8.9%, 6.3% and 5.9% better compared with MnoKG on CHIP-2020, CHIP-2022 and DiaKG datasets, respectively. It shows that using medical knowledge graph is important for providing external contexts.

(2) Our model outperforms **MEnt**. For example, the overall improvements are 1.5%, 1.6% and 0.5% compared with MEnt on CHIP-2020, CHIP-2022 and DiaKG datasets, respectively. It justifies that the second-order graph provides a positive feedback for the first-order representations and further enhance the encoding of entity mentions in medical text.

(3) Our model surpasses **MRel** and the rate is less than that of **MEnt**. It seems the second-order graph plays a major role in dual heterogeneous graph for joint extraction task.

(4) Our model precedes **MnoGNN**. Specifically, the exceeding values are 1.2%, 2.5% and 0.1% for the datasets. It shows that the GNN update part is vital for a better knowledge graph information fusion.

(5) The **MEnt** model only added the first-order graph, and the **MRel** only added the second-order graph. It can be seen from the data that the effect of **MRel** is better than that of **MEnt**. Second-order graph can improve entity extraction even more effectively, because it acts directly on the entity span.

(6) Our model performs better than **MRAtt**. To be specific, the overall improvement of our model is 0.5%/1.5% compared with MRAtt on CHIP-2020/CHIP-2022 datasets and the result is comparable to that of MRAtt on DiaKG dataset. It means that the selective gate mechanism can improve the performance for relation fusion.

5 Conclusions

In this paper, we study the problem of the joint entity and relation extraction in the medical text. We propose to construct the high-order heterogeneous graph from the medical knowledge graph, and learn the entity span representations in a knowledge-enhanced which integrates the deep representations from both global and local contexts. The extraction of entities and relations is in a collective way. The experimental results show that our model is more effective than state-of-the-art methods.

Limitations

We inject the medical knowledge graph into local texts for entity span representations enhancement. However, unlike most joint extraction methods, the proposed model is hard to be trained in a parallel way. Therefore, it is time-consuming to obtain a well-trained model. We would like to optimize the architecture of the model in the future.

Moreover, our model is adapted to Chinese medical texts where a token usually means a character (not a word). Hence, there will be errors when aligning the entities from the knowledge graph with mentions from local texts. Word segmentation task will be considered in our future work.

Acknowledgements

This work is supported by China Mobile Holistic Artificial Intelligence Major Project Funding (R22105ZS, R22105ZSC01) and National Key R&D Program of China (2021ZD0140408).

References

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çaglar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charlie Nash, Victoria Langston, Chris Dyer, Nicolas Manfred Otto Heess, Daan Wierstra, Pushmeet Kohli, Matthew M. Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. 2018. Relational inductive biases, deep learning, and graph networks. *ArXiv*, abs/1806.01261.

Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Adversarial training for multi-context joint entity and relation extraction. In *EMNLP*.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Dejie Chang, Mosha Chen, Chaozhen Liu, Liping Liu, Dongdong Li, Wei Li, Fei Kong, Bangchang Liu, Xiaobin Luo, Ji Qi, et al. 2021. Diakg: An annotated diabetes dataset for medical knowledge graph construction. In *China Conference on Knowledge Graph and Semantic Computing*, pages 308–314. Springer.

Markus Eberts and Adrian Ulges. 2020. Span-based joint entity and relation extraction with transformer pre-training. In *ECAI 2020*, pages 2006–2013. IOS Press.

Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. 2019. Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1409–1418.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 1263–1272.

T. Guan, H. Zan, X. Zhou, H. Xu, and K Zhang. 2020. *CMeIE: Construction and Evaluation of Chinese Medical Information Extraction Dataset*. Natural Language Processing and Chinese Computing, 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part I.

Kung-Hsiang Huang, Mu Yang, and Nanyun Peng. 2020. Biomedical event extraction with hierarchical knowledge graphs. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1277–1285, Online. Association for Computational Linguistics.

Bin Ji, Jie Yu, Shasha Li, Jun Ma, Qingbo Wu, Yusong Tan, and Huijun Liu. 2020. Span-based joint entity and relation extraction with attention-based span-specific and contextual semantic representations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 88–99.

Xiaodong Jiang, Ronghang Zhu, Sheng Li, and Pengsheng Ji. 2020. Co-embedding of nodes and edges with graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Tuan Lai, Heng Ji, ChengXiang Zhai, and Quan Hung Tran. 2021. Joint biomedical entity and relation extraction with knowledge-enhanced collective inference. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6248–6260, Online. Association for Computational Linguistics.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197.

Fei Li, Meishan Zhang, Guohong Fu, and Donghong Ji. 2017. A neural joint model for entity and relation extraction from biomedical text. *BMC bioinformatics*, 18(1):1–11.

Yang Li, Guodong Long, Tao Shen, Tianyi Zhou, Lina Yao, Huan Huo, and Jing Jiang. 2020. Self-attention enhanced selective gate with entity-aware embedding for distantly supervised relation extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8269–8276.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3219–3232.

Yi Luan, Dave Wadden, Luheng He, Amy Shah, Mari Ostendorf, and Hannaneh Hajishirzi. 2019. A general framework for information extraction using dynamic span graphs. In *Proceedings of NAACL-HLT*, pages 3036–3046.

Ling Luo, Zhihao Yang, Mingyu Cao, Lei Wang, Yin Zhang, and Hongfei Lin. 2020. A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature. *Journal of Biomedical Informatics*, 103:103384.

Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*.

Yali Pang, Tong Zhou, and Zhichang Zhang. 2021. A joint model for chinese medical entity and relation extraction based on graph convolutional networks. In *2021 3rd International Conference on Natural Language Processing (ICNLP)*, pages 119–124. IEEE.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Yong Shang, Yu Tian, Min Zhou, Tianshu Zhou, Kewei Lyu, Zhixiao Wang, Ran Xin, Tingbo Liang, Shiqiang Zhu, and Jingsong Li. 2021. Ehr-oriented knowledge graph system: Toward efficient utilization of non-used information buried in routine clinical practice. *IEEE Journal of Biomedical and Health Informatics*, 25(7):2463–2475.

Severine Verlinden, Klim Zaporojets, Johannes Deleu, Thomas Demeester, and Chris Develder. 2021. Injecting knowledge base information into end-to-end joint entity and relation extraction and coreference resolution. *arXiv preprint arXiv:2107.02286*.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1706–1721.

Shaolei Wang, Yue Zhang, Wanxiang Che, and Ting Liu. 2018. Joint extraction of entities and relations based on a novel graph scheme. In *IJCAI*, pages 4461–4467. Yokohama.

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1476–1488.

Kui Xue, Yangming Zhou, Zhiyuan Ma, Tong Ruan, Huanhuan Zhang, and Ping He. 2019. Fine-tuning bert for joint entity and relation extraction in chinese medical text. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 892–897. IEEE.

Zhiheng Yan, Chong Zhang, Jinlan Fu, Qi Zhang, and Zhongyu Wei. 2021. A partition filter network for joint entity and relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 185–197, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Shan Zhao, Minghao Hu, Zhiping Cai, and Fang Liu. 2021. Modeling dense cross-modal interactions for joint entity-relation extraction. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4032–4038.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *North American Association for Computational Linguistics (NAACL)*.

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
the "Limitations" section which is next to the section 5

☑ A2. Did you discuss any potential risks of your work?
the "Limitations" section which is next to the section 5

☑ A3. Do the abstract and introduction summarize the paper's main claims?
the 1st section (introduction) and the "abstract" section which is ahead of section 1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☒ Did you use or create scientific artifacts?

Left blank.

☐ B1. Did you cite the creators of artifacts you used?
No response.

☐ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
No response.

☐ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
No response.

☐ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
No response.

☐ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
No response.

☐ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
No response.

C ☑ Did you run computational experiments?

section 4 (experiments)

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
section 4.1.4 (Implementation details)

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
section 4.1.4 (Implementation details)

☒ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
the results are stable without large error bars

☒ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
not used

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.