Journal of Intelligent Information Systems
https://doi.org/10.1007/s10844-023-00826-9

RESEARCH

A mutually enhanced multi-scale relation-aware graph convolutional network for argument pair extraction

Xiaofei Zhu[1] · Yidan Liu[1] · Zhuo Chen[1] · Xu Chen[2] · Jiafeng Guo[3] · Stefan Dietze[4,5]

Received: 19 August 2023 / Revised: 24 October 2023 / Accepted: 25 October 2023

Abstract
Argument pair extraction (APE) is a fine-grained task of argument mining which aims to identify arguments offered by different participants in some discourse and detect interaction relationships between arguments from different participants. In recent years, many research efforts have been devoted to dealing with APE in a multi-task learning framework. Although these approaches have achieved encouraging results, they still face several challenging issues. First, different types of sentence relationships as well as different levels of information exchange among sentences are largely ignored. Second, they solely model interactions between argument pairs either in an explicit or implicit strategy, while neglecting the complementary effect of the two strategies. In this paper, we propose a novel Mutually Enhanced Multi-Scale Relation-Aware Graph Convolutional Network (MMR-GCN) for APE. Specifically, we first design a multi-scale relation-aware graph aggregation module to explicitly model the complex relationships between review and rebuttal passage sentences. In addition, we propose a mutually enhancement transformer module to implicitly and interactively enhance representations of review and rebuttal passage sentences. We experimentally validate MMR-GCN by comparing with the state-of-the-art APE methods. Experimental results show that it considerably outperforms all baseline methods, and the relative performance improvement of MMR-GCN over the best performing baseline MRC-APE in terms of F1 score reaches to 3.48% and 4.43% on the two benchmark datasets, respectively.

Keywords Argument mining · Argument pair extraction · Transformer · Graph convolutional network

1 Introduction

Argument Mining (AM) has attracted increasing research interest in recent years due to its wide and practical usage in many real-world scenarios, such as AI debate (Le et al., 2018; Mancini et al., 2022), social media (Lytos et al., 2019; Dutta et al., 2022), legal field (Westermann et al., 2022; Elaraby & Litman, 2022), etc. AM aims to extract the semantic and logical structure of argumentative documents (Yuan et al., 2021; Cheng et al., 2022). Early research works mainly focus on identifying the argumentation structure of a monological

Extended author information available on the last page of the article document (Potash et al., 2017; Kuribayashi et al., 2019). Potash et al. (Potash et al., 2017) employ pointer network sequence-to-sequence attention modeling to detect arguments and extract relationships between detected arguments. Kuribayashi et al. (2019)) further take into account linguistic features to predict argument structure.

Recently, there is a new research line of work investigating the extraction of interactive argument pairs from two argumentative passages of a discussion, which is referred to as argument pair extraction (APE) (Bao et al., 2021, 2022; Lu et al., 2021). APE is a challenging task since it needs to identify arguments offered by different participants in some discourse and detect interaction relationships between arguments from different participants. Figure 1 shows an example of APE, which comprises a review passage with eleven sentences and a rebuttal passage with sixteen sentences. The APE model needs to identify all arguments (marked in blue and brown) in both the review passage and the rebuttal passage, as well as extract all argument pairs (e.g., Pair-1 and Pair-2).

A straightforward solution is to cast the APE task into a pipeline with two subtasks, including a sequence labeling task which identifies all arguments from two passages (i.e., the review passage and the rebuttal passage), as well as a sentence relation classification task which determines argument pairs from these identified arguments. However, the limitation of this solution is that it neglects the correlation information among the two subtasks. To overcome the issue, some research efforts have been devoted to training the two subtasks simultaneously in a multi-task learning framework. For example, Cheng et al. (2020) treat this task as a sequence labeling subtask and sentence relation classification subtask, and jointly optimize the two subtasks. Cheng et al. (2021) propose a multi-layer multi-cross encoder to implicitly learn the useful information in the two passages. Bao et al. (2021) construct an inter-sentence relation graph based on co-occurring words to explicitly model the relationship between the review passage sentences and rebuttal passage sentences. Bao et al. (2022) further cast the APE task as a multi-turn machine reading comprehension (MRC) task with an AM phase and an APE phase. Although these researches have considerably boosted the performance of the APE task, they still suffer from two issues. First, the complex relationships of both intra-passage and inter-passage sentences are not well explored, such as the different types of sentence relationships as well as the different level of information exchange among sentences. Second, they mainly capture interactions between argument pairs either in an explicit or implicit strategy. Since the two strategies usually complement to each other, only relying on one of them would result in inferior performance.

To handle the above issues, in this paper, we propose a Mutually Enhanced Multi-Scale Relation-Aware Graph Convolutional Network (MMR-GCN) for APE. In particular, we first leverage an encoder to obtain sentence representations in the review passage and rebuttal passage, then we construct a multi-scale relation-aware graph to explicitly model rich relationship information between two passages. After that, we design a mutually enhancement transformer module to implicitly utilize the sentence relationship between two passages, which can avoid the issue that argument pairs are difficult to extract due to the absence of co-occurrence words. It is worth noting that the two modules can be stacked multiple times to better capture the complex sentence relationship. In summary, our major contributions are as follows:

- We design a multi-scale relation-aware graph to explicitly model the relationship between review and rebuttal passage sentences.
- To overcome the problem that argument pairs are difficult to extract due to the absence of co-occurrence words, we propose a mutually enhancement transformer module to implicitly and interactively enhance representations of both review and rebuttal passage sentences.

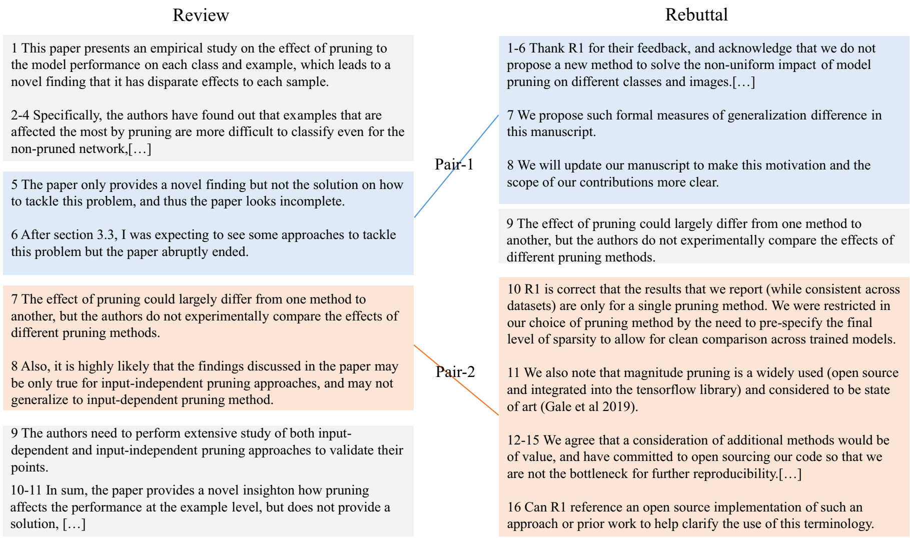


Fig. 1 An example of APE task

- Experimental results on two benchmark datasets (RR-Submission-v2 and RR-Passage) demonstrate that the proposed method significantly outperforms all baseline methods and achieves new state-of-the-art results.

2 Related work

Argument mining (AM) (Fromm et al., 2021; Srivastava et al., 2023; Andrea et al., 2021; Cabrio & Villata, 2022) attempts to extract the semantic and logical structure of argumentative documents, and plays a vital role in natural language processing (Le et al., 2018; Dutta et al., 2022; Elaraby & Litman, 2022; Yeginbergenova & Agerri, 2023). Previous works usually focus on modeling arguments in monological documents. Among these works, the neural network-based approaches have achieved promising results for AM. For example, Potash et al. (2017) predict argument structure by leveraging a sequence-to-sequence attention modeling, and also develop a joint model to simultaneously identify argument relationships. Eger et al. (2017) present a neural end-to-end solutions to AM, and propose to frame the AM problem as a dependency parsing problem. The major limitations of the two works is that they ignore the rich linguistic information in documents. To the end, Kuribayashi et al. (2019) propose to incorporate linguistic properties of documents, and explore a LSTM-minus-based span representation and a task-specific extended representation for AM.

Many recent research works attempt to analyze dialogical argumentation, and argument pair extraction (APE) has become an important research topic (Cheng et al., 2020). APE aims at extracting argument pairs from two passages of a discussion. Cheng et al. (2020) develop a multi-task learning framework which considers APE as two subtasks, i.e., a sequence labeling task and a sentence relation classification task. The former focuses on detecting arguments in both review and rebuttal passages, and the latter aims at identifying argument pairs between them. The two subtasks are then jointly optimized and the argument pairs are extracted by the combination of them. Its main drawback is that it simply concatenates the two passages to a single passage to perform the sequence labeling task, which ignores the different characteristics of the two passages. To tackle this issue, Cheng et al. (2021) apply two individual sequence encoders for the review passage or rebuttal passage respectively and update each encoder via mutual attention. They model the correlation between two passages via a table-filling model and extract argument pairs by leveraging the attention-guided multi-layer multi-cross encoding scheme.

More recently, some research efforts (Bao et al., 2021, 2022) have been devoted to modelling argument-level relations rather than only considering sentence-level relations. Bao et al. (2021) propose a mutual guidance framework. It uses the argument information in one passage to identify its paired arguments in the other passage. Bao et al. (2022) further cast the APE task as a multi-turn machine reading comprehension (MRC) task with an AM phase and an APE phase. Differ from these methods, we argue that the complex relationships of both intra-passage and inter-passage sentences should be explored. Since relational graph convolutional network can capture complex relationships between different data sources, and there have been many applications in the field of natural language processing (Liang et al., 2021; Xing & Tsang, 2022).Therefore, we develop a multi-scale relation-aware graph module to explicitly model the relationship between review passage sentences and rebuttal passage sentences. Moreover, we also propose to implicitly capture the sentence relationship between two passages by introducing a mutually enhancement transformer module, which can alleviate the relationship sparsity issue existing in the explicit relationship modelling strategy.

3 Problem definition

In the APE task, there are a review sequence $S^v = \{s_1^v, s_2^v, \cdots, s_m^v\}$ and a rebuttal sequence $S^b = \{s_1^b, s_2^b, \cdots, s_n^b\}$, where m and n denote the number of sentences in S^v and S^b, respectively. We treat the APE task as a multi-task learning problem, which consists of two subtasks, i.e., argument mining and sentence pairing. The former subtask aims at recognizing all arguments in each sequence which can be casted as a sentence-level sequence labeling problem. We use the standard BIO scheme (Ratinov & Roth, 2009) and Conditional Random Field (Lafferty et al., 2001) to extract all review arguments $A^v = \{a_1^v, a_2^v, \cdots, a_{l_v}^v\}$ and rebuttal arguments $A^b = \{a_1^b, a_2^b, \cdots, a_{l_b}^b\}$, where l_v and l_b denote the number of arguments in the review and rebuttal sequences, respectively. The latter subtask attempts to predict argument pairs from the review and rebuttal sequences, which can be formulated as a table-filling problem (authorname, 2014). We use the table-filling method and multi-layer perceptron to predict whether two sentences belong to the same argument pair. Finally, we get all the argument pairs $P = \{p_1, p_2, \cdots, p_{l_p}\}$ based on the two subtasks, where l_p denotes the number of argument pairs. As illustrated in Figure 1, there are two arguments in the review sequence (i.e., $A^v = \{(5-6), (7-8)\}$) and two arguments in the rebuttal sequence (i.e., $A^b = \{(1-8), (10-16)\}$). The argument pairs between the review arguments and rebuttal arguments are $P = \{((5-6), (1-8)), ((7-8), (10-16))\}$.

4 Methods

The overall architecture of our proposed model (MMR-GCN) is illustrated in Figure 2. It mainly consists of four components, including Sentence Encoder, Multi-Scale Relation-Aware Graph Aggregation (MSRAGA), Mutual Enhancement Transformer (MET), and Prediction Layer. First, each sentence in the review and rebuttal passages is fed into a sentence encoder to get a sentence representation. Next, a multi-scale relation-aware graph aggregation is used to explicitly model the relationship between review and rebuttal passage sentences. After that, we propose a mutual enhancement transformer to implicitly and interactively enhance the representations of review and rebuttal passage sentences. Note that, both MSRAGA and MET modules can be stacked multiple times. Finally, a prediction layer is used to extract arguments and identify argument pairs.

4.1 Sentence encoder

For each sentence $s_i = \{w^i_1, w^i_2, ..., w^i_{|s_i|}\}$ with $|s_i|$ words, we adopt the pre-trained language mdoel BERT (Devlin et al., 2019) to encode its content information and obtain the sentence's hidden state sequence $\{\mathbf{h}^i_1, \mathbf{h}^i_2, ..., \mathbf{h}^i_{|s_i|}\}$, then the hidden state sequence are fed into a word-level Bidirectional Long Short-Term Memory (Bi-LSTM) (Hochreiter & Schmidhuber, 1997). We concatenate the last hidden state of each direction as the sentence representation $\mathbf{s}_i$. Through the above operation, we can get the representation of all sentences in the review sequence $\mathbf{S}^v = \{\mathbf{s}^v_1, \mathbf{s}^v_2, ..., \mathbf{s}^v_m\}$ and the rebuttal sequence $\mathbf{S}^b = \{\mathbf{s}^b_1, \mathbf{s}^b_2, ..., \mathbf{s}^b_n\}$.

To further incorporate the contextual information, the obtained representations corresponding to the review and rebuttal sequences are fed into two distinct sentence-level Bi-LSTM to get the contextualized representations of the two sequences:

$$\mathbf{C}^v = \{\mathbf{c}^v_1, \mathbf{c}^v_2, ..., \mathbf{c}^v_m\}, \tag{1}$$

$$\mathbf{C}^b = \{\mathbf{c}^b_1, \mathbf{c}^b_2, ..., \mathbf{c}^b_n\}, \tag{2}$$

where $\mathbf{C}^v$ and $\mathbf{C}^b$ represent the review and rebuttal sequence representations, respectively.

4.2 Multi-Scale Relation-Aware Graph Aggregation (MSRAGA)

To effectively capture the rich relationship information embedded in both review and rebuttal sequences, we develop a multi-scale relation-aware graph aggregation (MSRAGA) module. MSRAGA consists of two sub-modules, i.e., Multi-scale Relation-aware Graph Construction and Graph Aggregation.

Multi-scale relation-aware graph construction Given a review sequence S^v and a rebuttal sequence S^b, we denote the graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. The nodes in $\mathcal{G}$ are the sentences in both review and rebuttal sequences, i.e., $\mathcal{V} = \mathcal{V}^v \cup \mathcal{V}^b$, where $\mathcal{V}^v = \{s^v_1, s^v_2, \cdots, s^v_m\}$ and $\mathcal{V}^b = \{s^b_1, s^b_2, \cdots, s^b_n\}$ are the node sets corresponding to the review sentences and the rebuttal sentences, respectively. The edge $(i, j, r_{ij}) \in \mathcal{E}$ denotes the information aggregation from the ith node to the jth node under the relation $r_{ij} \in \mathcal{R}$. In this paper, we define seven relationship types, i.e., $\mathcal{R} = \{r_1, r_2, \cdots, r_7\}$, and the definition of each relationship type is given in Table 1. More precisely, in the review sequence, we consider the sequential sentence

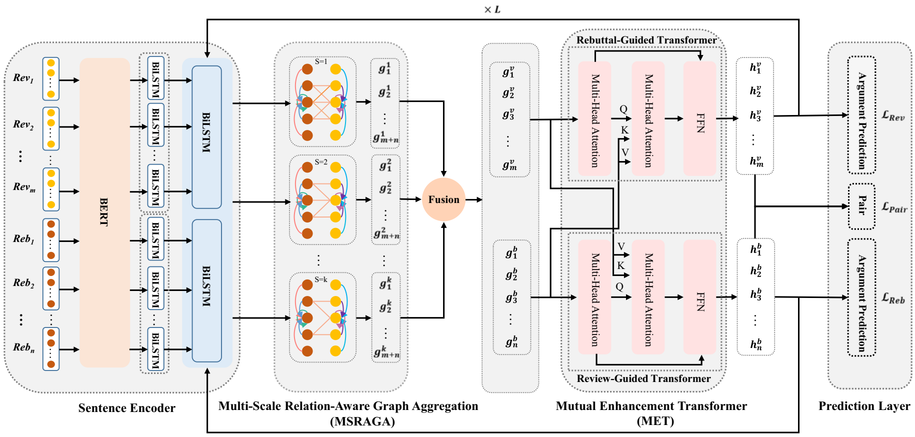


Fig. 2 The overview of our model structure

relationships and build edges between neighboring sentences. Since the distances between two neighboring sentences reflect their strength of relationships, we consider two different types of relationships, i.e., one-hop neighbor relationship and two-hop neighbor relationship. Moreover, we also introduce a self-loop relationship to hold the influence of the identical node. Similarly, we can obtain three corresponding relation types from the rebuttal sequence. At last, we also build cross-passage edges by exploiting the word co-occurrence relationship between the review sentences and the rebuttal sentences, i.e., there is a co-occurrence edge between a review sentence and a rebuttal sentence only when the number of co-occurrence words of the two sentences is greater than a certain threshold. Figure 3 demonstrates an example of graph construction, where different colors denote different types of edges.

Since a larger threshold usually leads to a more sparse connection between the review sentences and the rebuttal sentences, which means different thresholds will exploiting different levels of information exchange among the review and rebuttal passages. Therefore, in this work, we build multiple graphs with different scales of thresholds, e.g., $\mathcal{G}^s$ denotes the graph with a threshold s.

Graph aggregation For each node in $\mathcal{V}$, their feature vectors are initialized with corresponding contextualized sentence representations:

$$\mathbf{G} = \mathbf{C}^v \cup \mathbf{C}^b, \tag{3}$$

where $\mathbf{G} = \{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_{m+n}\} \in \mathbb{R}^{(m+n)\times 2d_h}$. Then we apply the relational graph convolution network (Schlichtkrull et al., 2018; Fang et al., 2023) to model the semantic interaction between sentences:

$$\mathbf{g}_i^s = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^{r,s}} \frac{1}{c_{i,r}}\mathbf{W}_r^s\mathbf{g}_j^s + \mathbf{W}_g^s\mathbf{g}_i^s\right), \tag{4}$$

where $\mathcal{N}_i^{r,s}$ denotes the neighbor set of node i under relation $r \in \mathcal{R}$ at scale s, and $c_{i,r}$ is a problem-specific normalization constant.

Afterwards, we apply a gate mechanism to fuse the representations of corresponding nodes at different scales to update sentence representation $\mathbf{g}_i$.

$$\mathbf{g}_i = \sum_{s \in S} \alpha^s \mathbf{g}_i^s, \tag{5}$$

$$\alpha^s = Softmax(\mathbf{W}^s \mathbf{g}_i^s + \mathbf{b}^s), \tag{6}$$

where $\mathbf{W}^s \in \mathbb{R}^{1 \times 2d_h}$ and $\mathbf{b}^s \in \mathbb{R}$ are trainable parameters, S denotes the set of thresholds. Hence, the review sequence and rebuttal sequence can be represented as $\mathbf{G}^v = \{\mathbf{g}_1^v, \mathbf{g}_2^v, \cdots, \mathbf{g}_m^v\}$ and $\mathbf{G}^b = \{\mathbf{g}_1^b, \mathbf{g}_2^b, \cdots, \mathbf{g}_n^b\}$, respectively.

4.3 Mutual Enhancement Transformer (MET)

In the multi-scale relation-aware graph aggregation module, we explicitly introduce the cross-passage relation, i.e., r_7, to enhance the representation learning of review (rebuttal) sentences by capturing their corresponding rebuttal (review) information. However, the cross-passage relation heavily relies on the co-occurrence words between a review sentence and a rebuttal sentence, which would suffer from the data sparsity issue, e.g., two correlated sentences would have no co-occurrence words. To handle this issue, we propose a mutual enhancement transformer to implicitly explore the semantic information among cross-passage sentences. The mutual enhancement transformer consists of two sub-modules, including a rebuttal-guided transformer and a review-guided transformer.

Rebuttal-guided transformer This module aims to use rebuttal semantic information to enhance review sequence representation, it consists of three parts, two multi-head attentions (Li et al., 2019) and a feed-forward neural network (FFN) (Rezatofighi et al., 2022). First, we adopt the multi-head self-attention mechanism to mine semantic information between review sentences. Specifically, for each attention head $i \in [1, h]$, we project review sequence embedding matrix $\mathbf{G}^v$ into the query, key, and value matrices, denoted as $\mathbf{Q}_i$, $\mathbf{K}_i$, $\mathbf{V}_i$, and concatenate the representation of each head as a representation of the updated review sequence:

$$\mathbf{U}^v = concat(\mathbf{U}_1^v, \mathbf{U}_2^v, \cdots, \mathbf{U}_h^v), \tag{7}$$

$$\mathbf{U}_i^v = Attention\left(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i\right), \tag{8}$$

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}})\mathbf{V}, \tag{9}$$

where h denotes the number of attention head, $Attention()$ is the attention function. For simplicity, we replace Equations (7) to (9) with the following equation:

$$\mathbf{U}^v = MultiHeadAttetnion(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \tag{10}$$

Second, we take the updated review sequence representation $\mathbf{U}^v$ as query, rebuttal sequence representation $\mathbf{G^b}$ as both key and value to enhance review sequence representation. This operation aims to reveal the importance of each review sentence from the perspective of rebuttal sentences, which do not rely on co-occurrence words between a review sentence and a rebuttal sentence. Since a review sequence may contain multiple arguments, it is more suitable to use multi-head attention to reveal the importance of a sentence from different perspectives.

$$\hat{\mathbf{U}}^v = MultiHeadAttention\left(\mathbf{U}^v, \mathbf{G}^b, \mathbf{G}^b\right) \tag{11}$$

Next, we apply FFN with a residual connection to generate the final review sequence representation $\mathbf{H}^v = \{\mathbf{h}_1^v, \mathbf{h}_2^v, \cdots, \mathbf{h}_m^v\}$, which contain information about itself and the rebuttal sequence:

$$\mathbf{H}^v = ReLU\left(\mathbf{W}\hat{\mathbf{U}}^v + \mathbf{b} + \mathbf{U}^v\right) \tag{12}$$

Review-guided transformer This module aims to use the review semantic information to enhance rebuttal sequence representation, it is similar to the rebuttal-guided transformer. A slight difference is that we take the rebuttal sequence as a query and the review sequence as both key and value of MultiHeadAttention to enhance the rebuttal sequence representation. Hence, we can obtain a new rebuttal sequence representation $\mathbf{H}^b = \{\mathbf{h}_1^b, \mathbf{h}_2^b, \cdots, \mathbf{h}_n^b\}$.

Note that, MSRAG and MEF modules are stacked L times. More specifically, we take $\mathbf{H}^v$ and $\mathbf{H}^b$ as input, then fed them into sentence-level Bi-LSTM again. In order to avoid the vanishing gradient problem, we also apply residual network to connect adjacent layers:

$$\mathbf{H}_{v/b}^{(l+1)} = LayerNorm\left(\mathbf{H}_{v/b}^{(l+1)} + \mathbf{H}_{v/b}^{(l)}\right) \tag{13}$$

4.4 Prediction layer

After new representations of the review sequence $\mathbf{H}^v$ and rebuttal sequence $\mathbf{H}^b$ are generated, we leverage them for argument mining and sentence pair prediction.

Argument mining For argument mining, we treat it a sequence labeling task and adopt CRF (Lafferty et al., 2001) to predict the argument label of each sentence. To be specific, given a review sequence $S^v = \{s_1^v, s_2^v, \cdots, s_m^v\}$, its corresponding argument label sequence

Table 1 All relationship types in the multi-scale relation-aware graph

$\mathcal{R}$	Definition
r_1	Self-loop relations of the review nodes
r_2	One-hop neighbor relations between the review nodes
r_3	Two-hop neighbor relations between the review nodes
r_4	Self-loop relations of the rebuttal nodes
r_5	One-hop neighbor relations between the rebuttal nodes
r_6	Two-hop neighbor relations between the rebuttal nodes
r_7	Cross relations between the review and rebuttal nodes

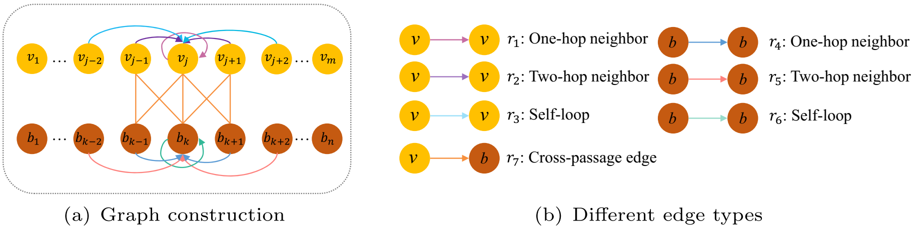


Fig. 3 An example of graph construction

$A^v = \{a_1^v, a_2^v, \cdots, a_m^v\}$, and $\mathcal{A}(S^v)$ denotes the set of possible argument label sequences for S^v. The probability of the review argument label sequence A^v is defined as follows:

$$P(A^v|S^v) = \frac{exp(score(S^v, A^v))}{\sum_{A' \in \mathcal{A}(S^v)} exp(score(S^v, A'))}, \tag{14}$$

where the score function $score(S^v, A^v)$ is the score for A^v which is the sum of transitions and emissions (Wang et al., 2020). Similarly, for a rebuttal sequence $S^b = \{s_1^b, s_2^b, \cdots, s_n^b\}$, we can define the probability of its argument label sequence $A^b = \{a_1^b, a_2^b, \cdots, a_n^b\}$ as $P(A^b|S^b)$. Then, the loss function for argument mining over both review sequence and rebuttal sequence is given by:

$$\mathcal{L}_{am} = -\left(\log P(A^v|S^v) + \log P(A^b|S^b)\right) \tag{15}$$

During inference, the predicted sequence label is defined as follows:

$$A_*^v = \arg\min_{A^v} P(A^v|S^v) \tag{16}$$

$$A_*^b = \arg\min_{A^b} P(A^b|S^b) \tag{17}$$

Sentence pair prediction We take $\mathbf{H}^v$ and $\mathbf{H}^b$ as the input of the sentence pair prediction module, and generate a table feature. Specifically, for a sentence pair (s_i^v, s_j^b), we denote it table feature as $\mathbf{TF}_{i,j}$, which is defined as follows:

$$\mathbf{TF}_{i,j} = \sigma\left(\mathbf{W}_p[\mathbf{h}_i^v;\mathbf{h}_j^b] + \mathbf{b}_p\right), \tag{18}$$

where $\mathbf{W}_p \in \mathbb{R}^{2d_h \times 4d_h}$ and $\mathbf{b}_p \in \mathbb{R}^{2d_h}$, are trainable parameters. $\mathbf{h}_i^v$ and $\mathbf{h}_j^b$ denote the representations of the i-th review sentence and j-th rebuttal sentence respectively, and σ is an activation function. After that, we employ a Multi-Layer Perceptron (MLP) and a sigmoid function to compute the probability $p_{i,j}$ that the two sentences (s_i^v, s_j^b) belong to the same argument pair.

For the sentence pair prediction module, we define the loss function $\mathcal{L}_{pair}$ as the cross-entropy of the prediction and the ground-truth:

$$\mathcal{L}_{pair} = -\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{i,j}^{pair} \log p_{i,j} + (1 - y_{i,j}^{pair}) \log(1 - p_{i,j})), \tag{19}$$

where $y_{i,j}^{pair}$ is the ground-truth which equals to 1 if s_i^v and s_j^b are paired and 0 otherwise. Finally, we obtain the overall loss of MRR-GCN:

$$\mathcal{L} = \mathcal{L}_{am} + \lambda \mathcal{L}_{pair} \tag{20}$$

where λ is a hyper-parameter to control the weight of two losses.

During inference, we consider a pair of candidate spans $([s_{i_1}^v, \cdots, s_{i_2}^v], [s_{j_1}^b, \cdots, s_{j_2}^b])$ as a valid pair if the two spans satisfy:

$$\frac{\sum_{i=i_1}^{i_2} \sum_{j=j_1}^{j_2} \mathbb{I}(p_{i,j} > 0.5)}{(i_2 - i_1 + 1) \times (j_2 - j_1 + 1)} \geq 0.5 \tag{21}$$

5 Experiments

5.1 Experimental setup

Datasets We evaluate our model on the Review-Rebuttal (RR) dataset (Cheng et al., 2020), which consists of 4,764 pairs of review-rebuttal passages collected from ICLR 2013-2020. Two versions of the dataset are provided, including RR-submission-v2 (Cheng et al., 2021) and RR-Passage. In each version, the data are split into train, development and test sets with a ratio of 8:1:1. Note that, in RR-submission-v2, multiple argument pairs of the same paper are in the same train/dev/test set, while RR-passage can not guarantee this. Table 2 presents the details about the RR dataset.

Baselines We compare MMR-GCN with the following competitive baselines:

- **MT-H-LSTM-CRF** Cheng et al. (2020) considers the APE task as two sub-tasks, and simultaneously optimizes the two sub-tasks with a multi-task learning framework.
- **MLMC** Cheng et al. (2021) leverages an attention-guided multi-layer multi-cross encoder and introduces a table-filling method for sentence relation classification.

Table 2 Statistics of the Review-Rebuttal(RR) dataset

RR	Total number of argument pairs	4764
Review	Total number of review sentences	99.8k
	Total number of review arguments	23.2k
	Total number of review argument sentences	58.5k
	Average number of sentences of per review passage	21.0
	Average number of sentences or per review argument	2.5
Rebuttal	Total number of rebuttal sentences	94.9k
	Total number of rebuttal arguments	17.7k
	Total number of rebuttal argument sentences	67.5k
	Average number of sentences of per rebuttal passage	19.9
	Average number of sentences or per rebuttal argument	3.8

- **MGF** Bao et al. (2021) proposes to explicitly model the relationship between two sentences by introducing a mutual guided framework with an inter-sentence relation graph.
- **MRC-APE** Bao et al. (2022) attempts to explore argument-level interactions and takes the APE task as a machine reading comprehension problem with two phases.

5.2 Implementation details and metrics

We implement our model MMR-GCN in PyTorch with a NVIDIA RTX 3090Ti GPU, and adopt the bert-base-cased as our based encoder.To optimize MMR-GCN, we utilize the AdamW optimizer (Loshchilov & Hutter, 2017) with a learning rate of 2e-4. The head number of the multi-head attention and the dropout rate are 4 and 0.5, respectively. The batch size is 1 and the epoch is 25. During the evaluation process, we select the best model parameters based on their performance on the development set,and then apply them to the test set to assess the final performance. We evaluate the compared methods using three metrics: Precision, Recall, and F1 scores. We report the performance of the argument pair extraction task, as well as the two sub-tasks, namely argument mining and sentence pairing. To ensure the reliability of our results, we conduct the experiments five times with different random states. We then calculate the average scores across these iterations and report them as the final evaluation results.

5.3 Overall comparison

The experimental results of MMR-GCN and all baseline methods on RR-Submission-v2 and RR-passage datasets are reported in Table 3. For the task of argument pair extraction, we can observe that MMR-GCN significantly improves the overall performance on both datasets as compared with all baseline methods. To be specific, among all baseline methods, MT-H-LSTM-CRF achieves the worst performance. This is attributed to that it could not fully capture the specific characteristics and relations of the review and rebuttal passages. MLMC obtains a better performance than MT-H-LSTM-CRF as it further models the sentence-level correlations between two passages by applying a table-filling method. MGF is superior to both MT-H-LSTM-CRF and MLMC. The reason is that it proposes to explicitly model the argument-level correlations between two passages with a mutual guidance framework. Similar to MGF, MRC-APE also models the argument-level interactions and achieves the best performance on both datasets in terms of F1 score among all baseline methods. This is mainly because that MRC-APE identifies all arguments based on an argument mining query, and extracts its paired arguments from another passage by taking each identified argument as an APE query. Our proposed method MMR-GCN considerably outperforms all baseline methods, and achieves the highest F1 score of 41.31% and 42.41% on RR-submisssion-v2 and RR-passage, respectively. The relative performance improvements of MMR-GCN over the best performing baseline MRC-APE in terms of F1 score reach to 3.48% and 4.43% on the two datasets, respectively.

In addition, for all methods, we also show their corresponding performance of the two sub-tasks, i.e., argument mining and sentence pairing. It is worth noting that both MGF and MRC-APE leverage the argument-level correlations rather than sentence-level correlations, therefore they cannot be utilized for the sentence pairing sub-task. For the

Table 3 Experimental results on RR-submisssion-v2 and RR-passage

Data	Models	Argument Mining			Sentence Pairing			Argument Pair Extraction		
		Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
RR-Submission-v2	**MT-H-LSTM-CRF** Cheng et al. (2020)	70.74	69.46	70.09	52.05	46.74	49.25	27.24	26.00	26.61
	MLMC Cheng et al. (2021)	69.53	**73.27**	71.35	60.01	46.82	52.60	37.15	29.38	32.81
	MGF Bao et al. (2021)	70.40	71.87	71.13	-	-	-	34.23	34.57	34.40
	MRC-APE Bao et al. (2022)	**71.83**	73.05	**72.43**	-	-	-	41.83	38.17	39.92
	MMR-GCN(Ours)	71.68	70.67	71.16	**68.85**	**59.81**	**63.99**	**44.69**	**38.44**	**41.31**
RR-Passage	**MT-H-LSTM-CRF** Cheng et al. (2020)	71.85	71.01	71.43	54.28	43.24	48.13	30.08	29.55	29.81
	MLMC(n=3) Cheng et al. (2021)	66.79	**72.17**	69.37	62.69	42.33	50.53	40.27	29.53	34.07
	MGF Bao et al. (2021)	73.62	70.88	72.22	-	-	-	38.03	35.68	36.82
	MRC-APE Bao et al. (2022)	**76.39**	70.62	**73.39**	-	-	-	37.70	**44.00**	40.61
	MMR-GCN(Ours)	71.47	71.60	71.54	**70.74**	**62.28**	**66.19**	**45.52**	39.70	**42.41**

The best results are in boldface, and “-” denotes that the corresponding baselines don’t treat APE as sequence labeling and sentence pairing task argument mining sub-task, we can see that our proposed approach MMR-GCN is relatively inferior to some baseline methods, such as MRC-APE. While for the sentence pairing sub-task, MMR-GCN is significantly superior to all baseline methods. Similar results can be observed for the other two metrics, i.e., precision and recall.

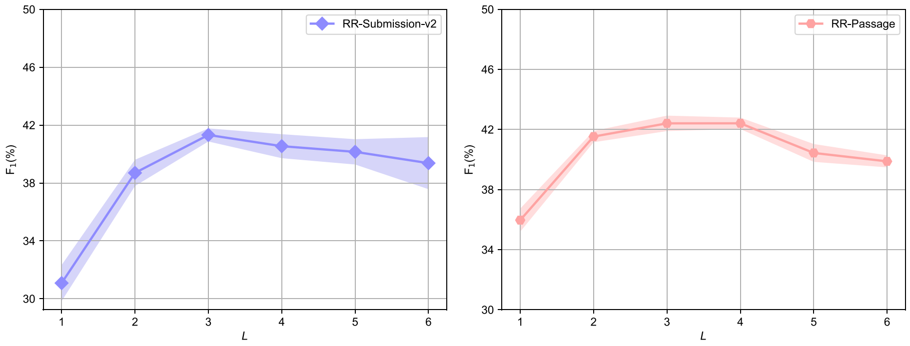
5.4 Ablation study

We conduct ablation studies on the RR-Submission-v2 dataset to analyze the impacts of different components in our model. More precisely, we consider the following variants for experiments: 1) *w/o MSRAGA*: we ignore the rich relationship information embedded in the review and rebuttal passage sentences by removing the multi-scale relation-aware graph aggregation module; 2) *w/o Scale-1*: we discard the relation-aware graph with scale 1, i.e., the number of co-occurrence words of a review passage sentence and a rebuttal passage sentence is greater than one; 3) *w/o Scale-2*: Similarly, we discard the relation-aware graph with scale 2; 4) *w/o MET*: we overlook the implicit semantic information among cross-passage sentence by removing the mutual enhancement transformer; 5) *w/o Relation*: All relationship types in the multi-scale relation-aware graph are discarded, and we replace the relational graph convolution network with a conventional graph convolution network.

Table 4 reports the performance of MMR-GCN and its variants on RR-Submission-V2 dataset. We summarize the observation as follows: Firstly, we can observe that removing each component will lead to a considerable performance drop. To be specific, the results show a performance decline after discarding the relation-aware graph with both scale 1 (*w/o Scale-1*) and scale 2 (*w/o Scale-2*), which indicates the effectiveness of employing multiple graphs with different scales in the proposed model. The results also show that removing the relation-aware graph with scale 2 will lead to a larger performance degradation as compared to removing the relation-aware graph with scale 1. The reason is that the graph with scale 1 would have more noisy edges, while the graph with scale 2 alleviate the noisy edge issue by utilizing a higher threshold. Secondly, removing the mutual enhancement transformer *w/o MET* or overlooking all relationship types in the multi-scale relation-aware graph *w/o Relation* will also cause a drop of performance. This indicates the importance of taking the relationship types among different passage sentences as well as modeling the implicit semantic information among cross-passage sentences. Third, among all variants, removing the multi-scale relation-aware graph aggregation module *w/o MSRAGA* will significantly degrade our model performance, e.g., the relative performance drop is 11.58%.

Table 4 Ablation study of our model on RR-Submission-V2 dataset

Model Settings	APE			
	Pre.	Rec.	F1	Δ F1
MMR-GCN	44.69	38.44	41.31	-
w/o MSRAGA	30.78	28.75	29.73	-11.58%
w/o Graph2	38.35	29.68	33.45	-7.86%
w/o Graph1	43.16	37.07	39.85	-1.46%
w/o MET	43.75	37.06	40.11	-1.20%
w/o Relation	43.28	37.35	40.07	-1.24%



(a) The impact of different L on the RR-Submission-v2 dataset.

(b) The impact of different L on the RR-Passage datast.

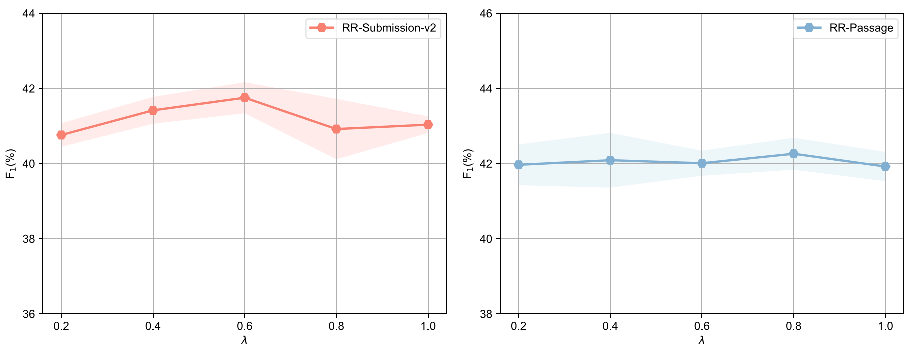
Fig. 4 The impact of stacked layers L

5.5 Hyperparameter sensitivity analysis

Here, we investigate the performance of MMR-GCN with respect to different settings of important hyperparameters.

Impact of L To investigate the impact of the stack times L of both MSRAG and MEF modules on the performance of MMR-GCN, we vary L from 1 to 6. The results are reported in Figure 4. On the RR-Submission-v2 dataset, we find that the performance of MMR-GCN first improves gradually and achieves the best results when L=3. There is a drop of performance when we continue to increase L. Similar trend of model performance can also be observed on the RR-Passage dataset. This implies that the parameter L is critical for boosting the performance of MMR-GCN. When L is small, the model cannot well capture the complex relationships between sentences of the two passages. However, when L increases to a certain extent (e.g., larger than 3), the model will suffer from the overfitting problem.

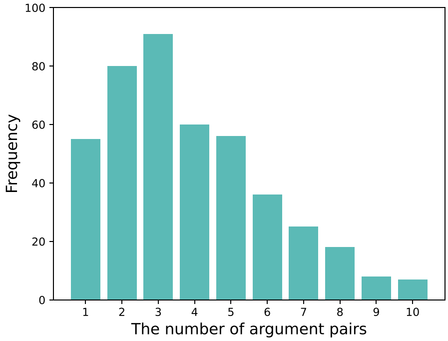
Impact of λ λ is a loss balancing parameter, which is utilized to adjust the weight of the two losses in our model MMR-GCN. We investigate the performance of MMR-GCN with λ varying from 0 to 1, and the results are shown in Figure 5. From Figure 5, we can see that



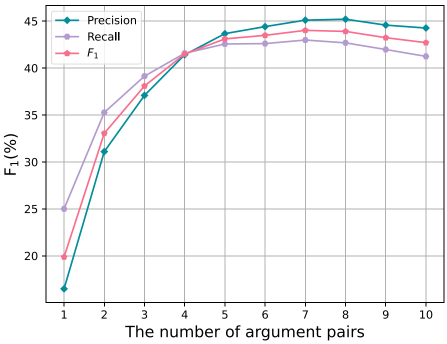
(a) The impact of different λ on the RR-Submission-v2 dataset.

(b) The impact of different λ on the RR-Passage dataset.

Fig. 5 The impact of hyperparameter λ



(a) Frequency distribution of argument pairs.



(b) F_1 score with different number of argument pairs.

Fig. 6 Analysis on different number of argument pairs

on the RR-Submission-v2 dataset, MMR-GCN achieves the best performance when λ is around 0.6, and the model performance has a drop when λ becomes too large or too small. The performance of MMR-GCN on the RR-Passage dataset is more stable as compared to that on the RR-Submission-v2 dataset, and the best performance is obtained when λ=0.8. This indicates that among the two sub-tasks, the argument mining is more important than the sentence pairing in our model since the error in the argument mining sub-task would be propagated into the sentence pairing sub-task.

5.6 Performance over different number of argument pairs

As different data samples may have distinct number of argument pairs, in this section, we investigate the performance of our proposed model MMR-GCN over samples with different argument pair numbers. As the number of argument pairs of most samples are between 1 and 10, we show their corresponding frequency distribution of argument pairs on the RR-Submission-v2 dataset in Figure 6(a). We can see that most samples have a relatively small number of argument pairs (e.g., less than 5), and samples with three argument pairs has the highest frequency. Figure 6(b) shows the performance of MMR-GCN over samples with different number of argument pairs. The results show that the performance raises gradually when samples have more argument pairs, and reaches a peak when the number of argument pairs is around seven. After that, there will be a performance degradation. This is because MMR-GCN leverages two modules (e.g., MSRAGA and MET) to explicitly and implicitly learn cross-passage relation information between review and rebuttal passages. When there is a small number of argument pairs or even no argument pairs, the cross relation information would be noise and hurt the model performance. However, when the number of argument pairs is too large, it will make the task become too complex and leads to inferior performance.

5.7 Computational efficiency

Table 5 shows the comparison results for computational efficiency between our model and MLMC, including the running time and number of parameters, on the RR-Submission-v2

Table 5 Running time per epoch and number of parameters of the proposed MMR-GCN and MLMC on the RR-Submission-v2 dataset

Models	RT(min)	#Params
MLMC(n=1)	22	5.8M
MLMC(n=2)	33	7.4M
MLMC(n=3)	44	9.1M
MLMC(n=4)	55	10.8M
MLMC(n=5)	66	12.5M
MMR-GCN(L = 1)	12	5.3M
MMR-GCN(L = 2)	13	6.6M
MMR-GCN(L = 3)	14	7.9M
MMR-GCN(L = 4)	15	9.2M
MMR-GCN(L = 5)	16	10.5M

dataset. Compared with MLMC, our model achieves better results in terms of both running time and number of parameters. Specifically, when the same number of layers is adopted, our proposed model MMR-GCN will run faster than MLMC. For example, the running time of MMR-GCN(L=2) is 13 minutes, which is nearly one third of the running time of MLMC(n=2). In addition, when we raise the number of layers, the improvement of the running time of MMR-GCN is relatively much small than that of MLMC. Finally, the number of parameters of MMR-GCN is also less than that of MLMC.

5.8 Impact of training set proportion

To provide more insights on the performance of our proposed model MMR-GCN with respect to the quality of training sets, we carry out comparative experiments under different training proportions on the dataset RR-Submission-v2. The results are demonstrated in Figure 7, from which we can have the following observations. First, the performance of MMR-GCN is sensitive to the size of the training set, and obtains superior performance when more training samples are available. Second, our proposed approach consistently performs better than the two most competitive baselines, i.e., MGF and MRC-APE.

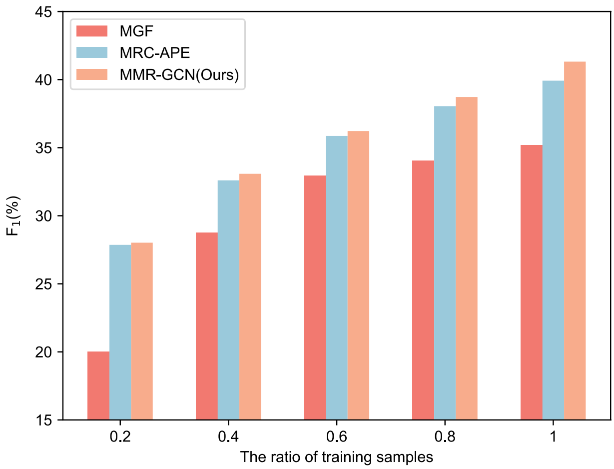
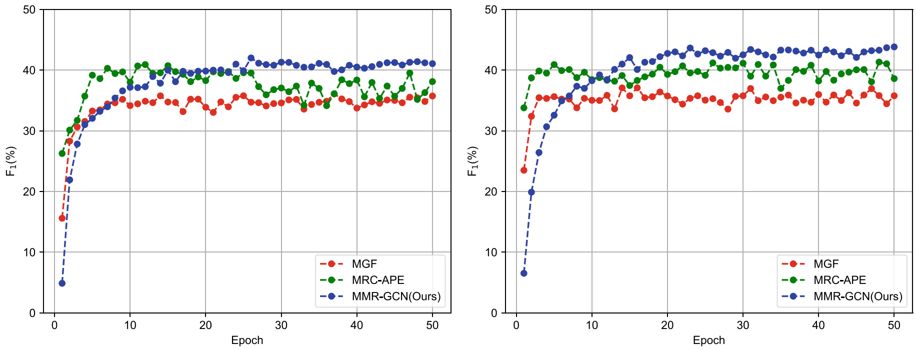


Fig. 7 Proportion of Training Data



(a) Learning curve on the RR-Submission-v2 dataset.

(b) Learning curve on the RR-Passage dataset.

Fig. 8 Learning curves of the proposed model MMR-GCN and the two most competitive baseline models (i.e., MGF and MRC-APE) on the two datasets (i.e., RR-Submission-v2 and RR-Passage)

5.9 Convergence analysis

Figure 8 demonstrates the convergence speed of our proposed model MMR-GCN and the two most competitive baseline models (i.e., MGF and MRC-APE) on the two datasets (i.e., RR-Submission-v2 and RR-Passage). The results show that MMR-GCN can obtain its best performance at a small epochs. For example, it achieves the best performance on both datasets at epoch 25, and after that it becomes stable and convergent. In addition, the performance of MMR-GCN is consistently better than that of both MGF and MRC-APE when the number of epochs is over 20 and 15 on RR-Submission-v2 and RR-Passage, respectively.

6 Conclusion

In this work, we present a novel mutually enhanced multi-scale relation-aware graph convolutional network (MMR-GCN) for the task of APE. To be specific, we propose to explore the complex relationships of both intra-passage and inter-passage sentences and develop a multi-scale relation-aware graph aggregation module to explicitly model different types of sentence relationships as well as the different levels of information exchange among sentences. In addition, we design a mutual enhancement transformer to implicitly explore the semantic information among cross-passage sentences, which can alleviate the data sparsity issue existing in the multi-scale relation-aware graph aggregation module. Experimental results show that the proposed model is superior to all baseline methods on all datasets.

For future work, we will exploit external knowledge to enhance the modeling of complex relationships among different sentences. It is worth noting that MMR-GCN achieves promising performance on the subtask of sentence pairing, while its performance on the subtask of argument mining is inferior to the best performing baseline. To the end, we also plan to improve the argument mining capability of our model.

Acknowledgements This work was supported by the National Natural Science Foundation of China [grant number 62141201]; the Natural Science Foundation of Chongqing, China [grant number CSTB2022NSCQ-MSX1672]; the Major Project of Science and Technology Research Program of Chongqing Education

Commission of China [grant number KJZD-M202201102]; the Federal Ministry of Education and Research [grant number 01IS21086].

Author contributions Xiaofei Zhu: Conceptualization, Methodology, Funding acquisition. Yidan Liu: Software, Methodology, Writing- Original draft preparation. Zhuo Chen: Writing- Reviewing and Editing. Xu Chen: Supervision, Funding acquisition. Jiafeng Guo: Supervision, Funding acquisition. Stefan Dietze: Writing- Reviewing and Editing.

Funding National Natural Science Foundation of China [grant number 62141201], Natural Science Foundation of Chongqing, China [grant number CSTB2022NSCQ-MSX1672], Major Project of Science and Technology Research Program of Chongqing Education Commission of China [grant number KJZD-M202201102], Federal Ministry of Education and Research [grant number 01IS21086].

Availability of supporting data RR-submission-v2 and RR-passage are open-source datasets and can be downloaded from https://github.com/ LiyingCheng95/ArgumentPairExtraction.

Declarations

Conflicts of interest The authors declare that there are no financial or non-financial interests directly or indirectly related to the work submitted for publication.

Ethical approval Not applicable.

References

Bao, J., Liang, B., Sun, J., et al. (2021). Argument pair extraction with mutual guidance and inter-sentence relation graph. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 3923–3934. https://doi.org/10.18653/v1/2021.emnlp-main.319

Bao, J., Sun, J., Zhu, Q., et al. (2022). Have my arguments been replied to? Argument Pair Extraction as Machine Reading Comprehension. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics,pp. 29–35. https://doi.org/10.18653/v1/2022.acl-short.4

Cabrio, E., & Villata, S. (2022). A data-informed analysis of argument mining. Journal of Logics and their Applications 9(4), 825–854. https://doi.org/10.24963/ijcai.2018/766

Cheng, L., Bing, L., He R., et al. (2022). IAM: a comprehensive and large-scale dataset for integrated argument mining tasks. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, pp. 2277–2287. https://doi.org/10.18653/v1/2022.acl-long.162

Cheng, L., Bing, L., Yu, Q., et al. (2020) APE: argument pair extraction from peer review and rebuttal via multi-task learning. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 7000–7011. https://doi.org/10.18653/v1/2020.emnlp-main.569

Cheng, L., et al.(2021) Argument pair extraction via attention-guided multi-layer multi-cross encoding. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, pp. 6341–6353. https://doi.org/10.18653/v1/2021.acl-long.496

Devlin, J., Chang, M.-W., Lee, K., et al. (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, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dutta, S., Juneja, J., Das, D., et al. (2022). Can Unsupervised Knowledge Transfer from Social Discussions Help Argument Mining? In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, pp. 7774–7786. https://doi.org/10.18653/v1/2022.acl-long.536

Eger, S., Daxenberger, J., & Gurevych I. (2017). Neural end-to-end learning for computational argumentation mining. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 11–22. https://doi.org/10.18653/v1/P17-1002

Elaraby, M., & Litman, D. J. (2022). ArgLegalSumm: improving abstractive summarization of legal documents with argument mining. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 6187–6194. https://doi.org/10.48550/arXiv.2209.01650

Fang Y., Li, X., Ye, R., et al. (2023) Relation-aware graph convolutional networks for multi-relational network alignment. ACM Transactions on Intelligent Systems and Technology *14*(2), 37:1–37:23. https://doi.org/10.1145/3579827

Fromm, M., Faerman, E., Berrendorf, M., et al. (2021). Argument Mining Driven Analysis of Peer-Reviews. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, pp. 4758–4766. https://doi.org/10.1609/aaai.v35i6.16607

Galassi, A., Lippi M., & Torroni P. (2021). Multi-task attentive residual networks for argument mining. arXiv:2102.12227. https://doi.org/10.1109/TASLP.2023.3275040

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Lu, J., Zhongyu W., Jing, L., et al. (2021). Discrete argument representation learning for interactive argument pair identification. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 5467–5478. https://doi.org/10.18653/v1/2021.naacl-main.431

Kuribayashi, T., Ouchi, H., Inoue, N., et al. (2019). An empirical study of span representations in argumentation structure parsing. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, pp. 4691–4698. https://doi.org/10.18653/v1/P19-1464

Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning, pp. 282–289

Le, D. T., Nguyen, C.-T., & Nguyen, K. A. (2018). Dave the debater: a retrieval-based and generative argumentative dialogue agent. In: Proceedings of the 5th Workshop on Argument Mining, pp. 121–130. https://doi.org/10.18653/v1/W18-5215

Li, J., Wang, L., Zhang, J., et al. (2019). Modeling intra-relation in math word problems with different functional multi-head attentions. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, pp. 6162–6167. https://doi.org/10.18653/v1/P19-1619

Liang, Y., Meng, F., Zhang, J., et al. (2021). A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis. Neuro Computing *454*, 291–302. ISSN: 0925-2312. https://doi.org/10.1016/j.neucom.2021.05.028

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. In: International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1711.05101

Lytos, A., Lagkas, T., Sarigiannidis, P. G., et al. (2019). The evolution of argumentation mining: From models to social media and emerging tools. Information Processing and Management *56*(6). https://doi.org/10.1016/j.ipm.2019.102055

Mancini, E., Ruggeri, F., Galassi, A., et al. (2022). Multimodal argument mining: a case study in political debates. In: Proceedings of the 9th Workshop on Argument Mining, pp. 158–170

Miwa, M., & Sasaki, Y. (2014). Modeling joint entity and relation extraction with table representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1858–1869. https://doi.org/10.3115/v1/D14-1200

Potash, P., Romanov, A., & Rumshisky, A. (2017). Here's my point: joint pointer architecture for argument mining. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1364–1373. https://doi.org/10.18653/v1/D17-1143

Ratinov, L., & Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning, pp. 147–155

Rezatofighi, H., Zhu, T., Kaskman, R., et al. (2022). Learn to Predict Sets Using Feed-Forward Neural Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence *44*(12), 9011–9025. https://doi.org/10.48550/arXiv.2001.11845

Schlichtkrull, M., Kipf, T. N., Bloem, P., et al. (2018). Modeling Relational Data with Graph Convolutional Networks. In: The Semantic Web, pp. 593–607. https://doi.org/10.1007/978-3-319-93417-4_38

Srivastava, P., Bhatnagar, P., & Goel A. (2023). Argument mining using bert and self-attention based embeddings. arXiv:2302.13906. https://doi.org/10.48550/arXiv.2302.13906

Wang, Y., Zhang, J., Ma, J., et al. (2020). Contextualized Emotion Recognition in Conversation as Sequence Tagging. In: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 186–195

Westermann, H., Savelka, J., Walker, V. R., et al. (2022). Toward an intelligent tutoring system for argument mining in legal texts. https://doi.org/10.48550/arXiv.2210.13635

Xing, B., & Tsang, I. (2022). DARER: dual-task temporal relational recurrent reasoning network for joint dialog sentiment classification and act recognition. In: Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland: Association for Computational Linguistics, May, 2022, pp. 3611–3621. https://doi.org/10.18653/v1/2022.findings-acl.286

Yeginbergenova, A., & Agerri, R. (2023). Cross-lingual argument mining in the medical domain. https://doi.org/10.48550/arXiv.2301.10527

Yuan, J., Wei, Z., Zhao, D., et al. (2021). Leveraging argumentation knowledge graph for interactive argument pair identification. In: Findings of the Association for Computational Linguistics, pp. 2310–2319. https://doi.org/10.18653/v1/2021.findings-acl.203

Authors and Affiliations

Xiaofei Zhu[1] · Yidan Liu[1] · Zhuo Chen[1] · Xu Chen[2] · Jiafeng Guo[3] · Stefan Dietze[4,5]

✉ Xu Chen
cx@cqut.edu.cn

Xiaofei Zhu
zxf@cqut.edu.cn

Yidan Liu
lyd@stu.cqut.edu.cn

Zhuo Chen
chenzhuo@cqut.edu.cn

Jiafeng Guo
guojiafeng@ict.ac.cn

Stefan Dietze
stefan.dietze@gesis.org

1 College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

2 College of Accounting, Chongqing University of Technology, Chongqing 400054, China

3 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

4 Knowledge Technologies for the Social Sciences, Leibniz Institute for the Social Sciences, Cologne 50667, Germany

5 Institute of Computer Science, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany