Leveraging Argumentation Knowledge Graph for Interactive Argument Pair Identification

Jian Yuan[1], Zhongyu Wei[1,2]*, Donghua Zhao[3], Qi Zhang[4], Changjian Jiang[5]*

[1]School of Data Science, Fudan University, China
[2]Research Institute of Intelligent and Complex Systems, Fudan University, China
[3]School of Mathematical Sciences, Fudan University, China
[4]School of Computer Science, Fudan University, China
[5]School of International Relations & Public Affairs, Fudan University, China
`{19210980107,zywei,dhzhao,qz,Changjian}@fudan.edu.cn`

Abstract

Interactive argument pair identification is essential in the context of dialogical argumentation mining. Existing research treats it as a problem of sentence matching and largely relies on textual information to compute the similarities. However, the interaction of opinions usually involves the background of the topic and requires reasoning of knowledge, which is beyond textual information. In this paper, we propose to leverage external knowledge to enhance the identification of interactive argument pairs. We construct the argumentation knowledge graph from the discussion thread of the target topic in the online forum. The interaction between the original argument and the reply is then represented as the path of concepts in the knowledge graph. In practice, we utilize Graph Convolutional Network (GCN) to learn the concept representation in the knowledge graph and use a Transformer-based encoder to learn the representation of paths. Finally, an information alignment network is employed to capture the interaction of textual information of conceptual information (both entity-level and path-level). Experiment results indicate that our model achieves state-of-the-art performance in the benchmark dataset. Further analysis demonstrates the effectiveness of our model for enforcing knowledge reasoning through paths in the knowledge graph.

1 Introduction

Argumentation Mining aims at analyzing the semantic and logical structure of argumentative texts. Existing research covers argument structure prediction (Morio et al., 2020; Li et al., 2020), persuasiveness evaluation (Al Khatib et al., 2020; El Baff et al., 2020) and argument summarization (Bar-Haim et al., 2020b,a). Most of them focus on mono-
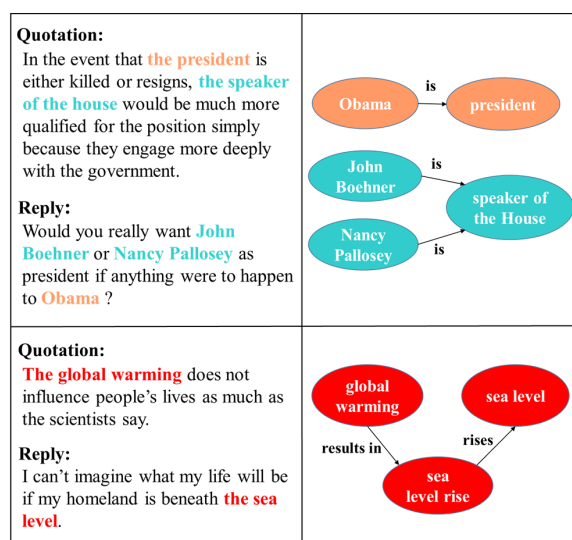
---

*Corresponding author.

Figure 1: Two instances of interactive argument pairs, the related concepts are colored same, and the corresponding knowledge is visualised in the right side.

logical context like student essays, public speeches, etc., where only one participant is involved.

Online forums such as idebate[1] and changemyview[2], enable people to exchange opinions on some specific topics freely. The user generated dataset of interactive arguments also motivates another line of research for argumentation in dialogical context (Asterhan and Schwarz, 2007). Initial researches in this filed focused on analyzing the ChangeMyView data (Tan et al., 2016; Wei et al., 2016) to summarize the key factors of persuasive arguments. Furthermore, Ji et al. (2019) and Cheng et al. (2020) propose the task of identifying and extracting interactive arguments. Ji et al. (2019) formulate this task as a problem of sentence pair scoring and computes the textual similarity

---

[1]https://idebate.org/
[2]https://www.reddit.com/r/changemyview/

between the two arguments as the result. Such task is then further applied to other fields such as legal domain. For instance, Yuan et al. (2021) organize a challenge aimed to identify the interactive arguments from the plaintiff and the defense in a legal case. However, the interaction of argumentation is beyond text matching.

Two sample pairs of interactive arguments are shown in Figure 1. Both pairs of arguments share a limited number of overlapping tokens and fail existing models. We have two observations. Firstly, background knowledge needs to be involved. In the first sample, we need to know that "Obama" is the "president", and both "John Boehner" and "Nancy Pallosey" are the "speaker of the house" to understand the context. Secondly, knowledge reasoning is necessary. In the second sample, the relationship between "global warming" and "sea level" is implied by a series of causal effects. Furthermore, as is shown in the example, an effective way of leveraging commonsense and causal effect knowledge is to find the reasoning paths between the concept entity pairs. Therefore, we argue that retrieving and understanding the reasoning paths should be incorporated for the identification of interactive arguments.

In this paper, we propose to leverage external knowledge to enhance the automatic identification of interactive arguments via background knowledge modeling and reasoning. We start with constructing an argumentation knowledge graph following (Khatib et al., 2020) based on the context of the discussion. Then, we extract entities of each argument and link them with the external knowledge graph to obtain the concept embedding as background knowledge. Besides, we generate paths connecting each pair of entities and encode them via a transformer encoder to enforce the reasoning. Finally, we integrate the entity embeddings, path representations, and textual embedding via an information alignment network to learn the final representation of the argument pair and output a real value as the matching score. We evaluate our proposed model on a publicly available dataset and experimental results show its effectiveness compared to some state-of-the-art approaches. Further analysis of the path encoding module reveals that our model is able to perform knowledge reasoning to some extent.

Statistics	w/o. grounding	w. grounding
# of nodes	291,199	291,199
# of edges	785,036	859,534
avg. degree	2.696	2.952
# of connected components	13,805	10,035

Table 1: Basic statistics on the argumentation graph without and with concept grounding.

2 Argumentation Knowledge Graph Construction

Data Source   The experimental dataset (Ji et al., 2019) in our research is constructed on top of the *CMV* dataset (Tan et al., 2016). In order to provide external knowledge for the identification of interactive arguments, we construct an argumentation knowledge graph based on the CMV dataset. *ChangeMyView* (CMV) is an online forum where users can either submit a post to elaborate their own viewpoints and invite other users to convince them of the opposite opinion or reply to others' posts to change the poster's original view. Tan et al. (2016) crawled 20,626 discussion threads with more than two posts from January 2013 to September 2015. We first extract all the concept-relation-concept triples (e_i^h, r_i, e_i^t) in the ith entry of the data source using Open Information Extraction (OpenIE). Our raw graph is thus $G = (V, E)$, where $V = \{e_i^h\}_{i=1}^n \cup \{e_i^t\}_{i=1}^n$ and $E = \{r_i\}_{i=1}^n$. The raw knowledge graph contains 291,199 nodes and 785,036 edges.

Concept Grounding   In order to further improve the quality of the knowledge graph, we conduct concept grounding to align all the nodes that share common conceptual meanings. Specifically, we use *WordNet* and Wikipedia API *TagMe* (Ferragina and Scaiella, 2010) in this process. If two concepts e_i, e_j are synonyms or refer to the same entry on Wikipedia, we add a new edge r_{equal} to the graph's edge set E. After concept grounding, the size of the edge set E expands to 859,534 with the size of the node-set V remained fixed. Some basic statistics of the knowledge graph are as shown in Table 1. It indicates that concept grounding increases the number of edges in a large margin and alleviates the problem of the sparsity of the original graph.

3 Proposed Model

Given an original argument q and its context c_q, and five candidate replies $\{r_i\}_{i=1}^5$ with their corresponding contexts $\{c_i\}_{i=1}^5$, the model needs to
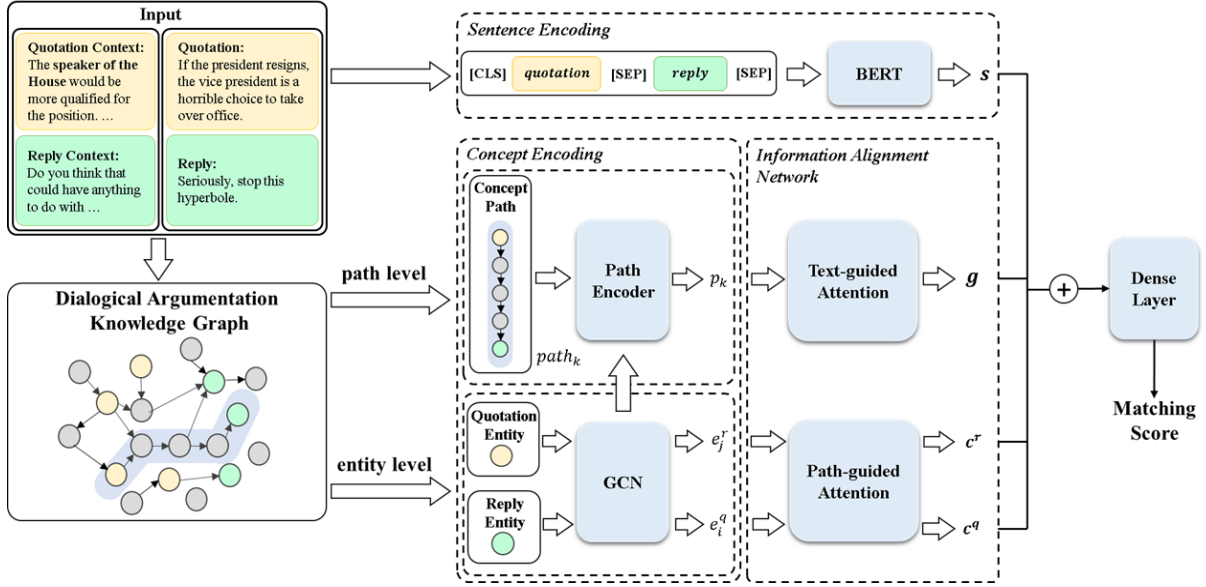
Figure 2: Illustration of the detailed architecture of our model to generate the matching feature vector, which mainly consists of three modules, a Sentence Encoder, a Concept Encoder and an Information Alignment Network. The output of these modules is then fed to a 2-layer perceptron to achieve the final matching score for the given argument pair.

identify the correct reply for q. We score each candidate pair independently and choose the reply with the highest score as the output. Moreover, in order to enable our model to conduct a reasoning process, we extract all the concept entities mentioned in the contexts from both sides, and also the concept paths that connect them. For simplicity, we will use sentence pairs to refer to the quotation and reply arguments and use concept pairs to refer to both the entities and paths in the following sections.

The full architecture of our scoring model is shown in Figure 2. It takes a sentence pair, and the concept pair extracted from its corresponding contexts as inputs and outputs a real value as its matching score. Our model mainly consists of three components, namely, sentence encoding, concept encoding, and information alignment network. We use a pre-trained language model, BERT, to learn the argument pair representation (§3.1), and encode the concept information from two levels, both entity level and path level with graph networks (§3.2). The information alignment network then integrates the sentence pair encoding and the concept encoding through a hierarchical attention mechanism to obtain the full matching features (§3.3), which are finally fed into a multi-layer perceptron (MLP) to calculate the final matching score (§3.4).

3.1 Sentence Encoding

As for the quotation and reply arguments, it is critical to use the semantic information implied in the texts. Various works have already proved the outstanding performance of pre-trained models in semantic modeling. In our work, we use the BERT model to generate the encoding s for the given argument pair by simply creating a sentence that takes the form of "[CLS] q [SEP] r [SEP]" and taking the embedding for the "[CLS]" token, just as suggested by previous works (Talmor et al., 2019).

3.2 Concept Encoding

For entities in the argumentation knowledge graph, we need to obtain the representation for each node. We use the BERT model with average pooling to get the initial representation for each entity. Then we encode the conceptual information in both entity-level and path-level with graph networks to enforce the background knowledge modeling and reasoning.

3.2.1 Entity Level Representation

To utilize the structural information entailed in the knowledge graph, we apply a 2-layer Graph Convolutional Network (GCN) to it. Here we adopt GCN as it has proved to be both effective and efficient in merging the node's neighbours' information into itself (Zhang et al., 2018).

2312

Formally, let $X \in \mathbb{R}^{n \times d}$ representing the embedding matrix for all n nodes, where each node's embedding is of size d. Denote $D \in \mathbb{R}^{n \times n}$ as the diagonal degree matrix and $A \in \mathbb{R}^{n \times n}$ as the adjacency matrix of the graph G. Then the normalized symmetric adjacency matrix of the graph G can be calculated as:

$$\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{1}$$

By feeding the graph G into the 2-layer GCN, the final graph representation $L \in \mathbb{R}^{n \times d}$ can be calculated as:

$$L = \sigma(\tilde{A} \sigma(\tilde{A} X W_0) W_1), \tag{2}$$

where σ stands for the non-linear function (RELU), and $W_0, W_1 \in \mathbb{R}^{d \times n}$ are trainable parameters of the network.

3.2.2 Path Level Representation

To further utilize the external knowledge, we want to encode the concept path retrieved from the knowledge graph, where a path starts from a concept e_q mentioned in the original post(i.e. quotation) q, traverses through the neighbored concepts, and finally ends at a concept e_r extracted from the reply r. For each concept pair, we choose the shortest path (if exists) between them as the path connecting them.

We use the GCN output as the representation for each node that appears in the path, hence we can denote the path between the i-th concept in q (c_i^q) and the j-th concept (c_j^r) in r as $P_{ij} = (c_i^q, c_1, ... c_{m_{ij}-1}, c_j^r) \in \mathbb{R}^{m_{ij} \times d}$, where m_{ij} is the the length of the path P_{ij}.

Transformer (Vaswani et al., 2017) has been shown powerful due to its self-attention mechanism, thus, we choose it to encode the path we collected from the knowledge graph. To underline the influence of the sequence in each path, we add the path's embedding with positional embedding PE. To sum up, our path encoder generates the representation for each $Path_{ij}$ as:

$$p_{ij} = \text{Transformer\_Encoder}(P_{ij} + PE). \tag{3}$$

The output is finally fed to a fully-connected layer to fit into the size of d.

3.3 Information Alignment Network

We then align the semantic information and the conceptual information through a hierarchical attention mechanism, i.e. a text-guided attention network for paths and a path-guided attention network for entities.

3.3.1 Text-guided Attention over Paths

Note that for the given argument pair and their contexts, we already have all the paths' encoding from previous modules. We first use attention between the k-th paths p_k and semantic vector s to integrate the encoding for all the paths g:

$$\alpha_k = s W_2 p_k, \tag{4}$$
$$\hat{\alpha} = \text{SoftMax}(\alpha), \tag{5}$$
$$g = \sum_k \hat{\alpha}_k p_k, \tag{6}$$

where W_2 is a parameter matrix to be learned, α and $\hat{\alpha}$ stands for the unnormalized and normalized attention weights.

3.3.2 Path-guided Attention over Concepts

Obtaining the full paths' representation g, we can further aggregate all the concepts' encoding $\{e_i\}$ of both sides using attention between them and g to generate the final representation of concepts c^q and c^r:

$$\beta_i^s = g W_3^s e_i^s, \tag{7}$$
$$\hat{\beta}^s = \text{SoftMax}(\beta^s), \tag{8}$$
$$c^s = \sum_i \hat{\beta}_i e_i^s, \tag{9}$$

where the subscript $s \in \{q, r\}$ indicates whether the concepts are from quotation or reply, W_3^q and W_3^r are parameters matrix to be learned, while $\hat{\beta}^q$ and $\hat{\beta}^r$ stand for the attention weights.

3.4 Matching Scoring

Eventually, we concatenate the textual information s, the reasoning paths information g and the concepts information c^q, c^r as the final feature and feed it to 2-layer perceptron to generate the matching score S of the given argument pair:

$$f = [s; g; c^q; c^r], \tag{10}$$
$$S = \sigma(W_S f + b_S), \tag{11}$$

where σ refers to the the rectified linear activation function (ReLU), W_S and b_S represent the weight vector and the bias respectively.

After obtaining the matching score for each argument pair, we treat the task as a sentence pair ranking problem, and use MarginRankingLoss for training:

$$L = \sum_{i=1}^{4} \max(0, \gamma - S(q, r^+) + S(q, r_i^-)), \tag{12}$$

where $S(q, r^+)$ refers to the matching score of the positive argument pair while $S(q, r_i^-)$ refers to the matching score of the i-th negative argument pair, and γ is the margin hyperparameter.

4 Experiments

In this section, we will introduce the dataset, the evaluation metrics, comparative models and experiment results.

4.1 Experiment Setup

Experimental Dataset We use the dataset constructed in (Ji et al., 2019) for evaluation. The authors find that in the *ChangeMyView* dataset (Tan et al., 2016), there exist replies that quote sentences from the original post. They extract all these *quotation-reply* pair q, r from posts in *Change-MyView* dataset (Tan et al., 2016). For every interactive argument pair, they randomly sample four negative replies $\{r_i^{neg}\}_{i=1}^4$ along with their contexts $\{c_i^{neg}\}_{i=1}^4$ from the same discussion thread. It contains 11,565 and 1,481 instances in training set and test set respectively. Furthermore, we randomly split 10% of the training set as validation set.

Implementation Details The output dimensions for the two layers in GCN are 256 and 128 respectively, the path transformer encoder we use is stacked by 6 encoder layers. The margin γ used in MarginRankingLoss is set to 0.5. Dropout is used as 0.1 to avoid overfitting. We use Adam as our optimizer with a learning rate set to 5×10^{-6} and weight decay set to 5×10^{-6}. We run our model for 100 epochs with early stop (Caruana et al., 2000).

Models for comparison We compare the performance of some state-of-the-art models.

- **BiGRU**: This method uses a Bidirectional GRU to encode the quotation and the reply argument separately and integrates their representations into a multilayer perceptron (MLP) to get the matching score.

- **VAE** : This method uses variational auto encoder through an encoder-decoder based architecture to get the encoding of the arguments and utilizes MLP for scoring.

- **DVAE** (Rolfe, 2017): This method substitutes the above *VAE* module with a discrete variational auto encoder and adopts the former framework.

- **BERT** (Devlin et al., 2019): This method fine-tunes the pre-trained BERT model for sentence-pair classification. Note that this model is not only a baseline model but also a sub-module of our proposed model.

Note that the above models only utilize the sentences of q and r, we also extend these models to incorporate context information.

- **RNN Context**: This method uses another Bi-GRU module to encode the context information of each argument and concatenate it with the argument representation to get the final features.

- **Hierarchical Context** (Ji et al., 2019): This method uses a token-level CNN with an attention mechanism to achieve the sentence-level information and then integrates such sentence representation with a BiGRU layer to obtain the final context encoding.

4.2 Overall Performance

We report both precision at one(P@1) and mean reciprocal rank (*MRR*) for evaluation. The performance of all the baseline models and our proposed model is as listed in Table 2. We have the following findings.

- Among all the context-agnostic baseline models, the BERT model achieves the highest performance, and it even defeats all other models that utilize the context information, indicating that such pre-trained language model does better encode the semantic information entailed in the texts.

- Incorporating context information is crucial for identifying interactive argument pairs, as is proved by the fact that all the context-aware models significantly outperform their counterpart baseline models.

- In comparison with all the context encoding methods, hierarchical context modeling outperforms the *RNN* method. Our method outperforms the hierarchical method, which proves the effectiveness of our model.

4.3 Ablation Study

The results of the ablation study are shown in the Table 3. We can find that the Path Transformer contributes greatly to the whole framework since

Methods	P@1(%)		MRR(%)	
	Dev Set	Test Set	Dev Set	Test Set
Random Guess	20	20	45.67	45.67
BiGRU	65.92	51.52	75.22	70.57
BiGRU+RNN Context	69.29	55.98	80.51	73.20
BiGRU+Hierarchical Context	70.93	57.46	82.47	73.72
VAE+Hierarchical Context	71.28	58.61	83.82	74.66
DVAE+Hierarchical Context*	73.70	61.17	85.14	76.16
BERT	73.18	61.85	84.69	76.57
BERT+Hierarchical Context	76.81	66.85	86.38	78.51
Ours	78.33	**68.75**	87.43	**80.85**

Table 2: Performance comparison for all the models on the development dataset and the test dataset, where the sign '*' represents the former state-of-the-art model. The best result on the test set is in **bold**

Model	P@1	MRR
Sentence + Concept + Align	68.75	80.85
: BiGRU as Path Encoder	67.12	79.46
: w/o Alignment Layer	65.48	79.39
: w/o Path Transformer	64.41	77.41
: w/o Concept Encoder	61.85	76.57
: w/o Sentence Encoder	51.96	68.83

Table 3: Ablation study on our proposed framework.

the model's performance drops by over 4% in P@1. This shows that besides the textual features and the concepts that directly appear in the argument, the concepts that emerge in the reasoning path are also important when considering whether two arguments have interactive relations.

5 Further Analyses

We conduct some further analyses to have a deeper understanding of the working mechanism of concept paths. Besides, we present an error analysis and a case study.

5.1 Analysis on Reasoning Path

Without ambiguity, we use positive paths and negative paths to refer to the paths that connecting the concepts in positive argument pair samples and the ones in the negative samples.

Connectivity between concept pairs First, we calculate the connectivity between each concept pair, and the results are as shown in Figure 3(a). In all the concept pairs of the two sides, the ones from a positive argument pair have a probability of 54% to form a reasoning path while the ones from a negative sample only have 41%, which conforms with

the fact that an interactive argument pair mainly talk about the same topic or subject.

Path length distribution We present the distribution of the length of concept paths in Figure 3(b). The vast majority of the path lengths lie in the range from 3 to 4, while the lengths of the positive paths are generally shorter than those of the negative paths. We owe such a difference in the average length to the fact that, in positive argument pairs, the replies tend to directly use the concepts mentioned in the quotation (the reasoning path between such concept pair is hence 1). Another interpretation for the shorter average length in positive pairs is that the longer the reasoning path is, the more likely it is to become an off-topic or off-subject reply.

Path relations We further analyze the types of concept paths generated by investigating the relations appearing in the path. As discussed in the Introduction(§1), there are mainly two types of external knowledge needed to handle the interactive argument pair identification task, namely the commonsense knowledge and the causal effect knowledge. Hence, we would like to see how much proportion of these two types of knowledge occurs in the reasoning paths respectively. For common sense knowledge, we pick out the relations that contain the **be verbs** and their variants, assuming such words indicate the relation of equivalence. And as for the causal effect relations, we use a set of lexical indicators from *+/-EffectWordNet* (Choi and Wiebe, 2014), *ConnotationWordNet* (Kang et al., 2014) and *Connotation Frames* (Rashkin et al., 2016). These lexicons all evaluate the causal effect of a given predicate. The distribution of relations under
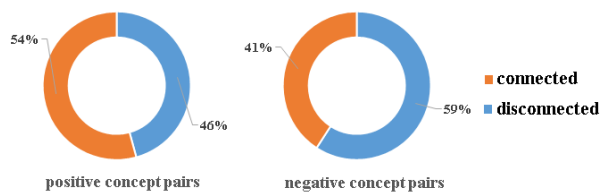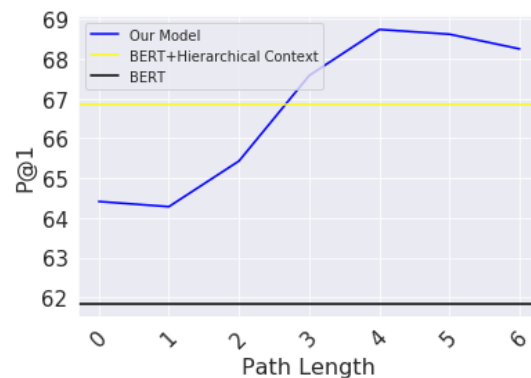
(a) The existence of reasoning paths between every concept pair in the contexts of the quotation and the reply.
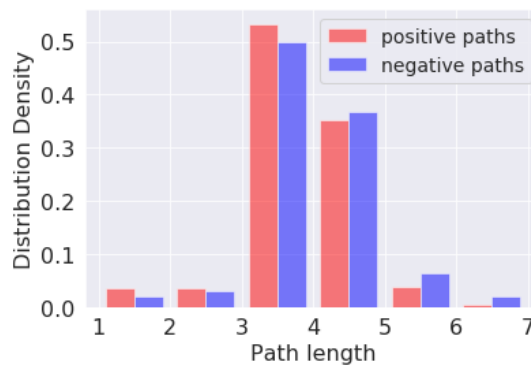
(b) The distribution of path length in the positive argument pair samples and in the negative samples.
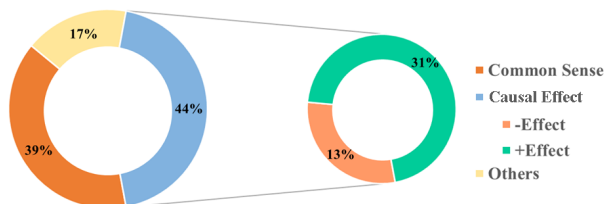
(c) The distribution of path relations in all the reasoning paths.

(d) The impact of different paths filtered by the path length on *P@1*.

Figure 3: The results of our further analysis on the reasoning paths.

such criterion is as shown in the Figure 3(c), from which we can find that near 40% of the relations belong to the common sense knowledge while 44% of the relations are of causal effect relations (31% for the positive effect and 13% for the negative effect).

Impact of path length on model performance We show the influence of the length of the path on the performance of the model in Figure 3(d). We set a threshold on the length of the path to filter concept paths used in the model. From the results, we can find that our model's performance improves significantly when the threshold is set from 3 to 4, in which most path lengths are distributed. The performance decreases when the path length is set to 5 and 6, which means it includes some noise and hurts the performance.

5.2 Error Analysis

For the instances that our model fails to predict the interactivity, we find that the problems are mainly

two-fold:

- **Concept level**: For some of the failed cases, we find that around 37% of them contain at least one reply from which no concepts can be extracted, which blocks our path-finding based reasoning process. It is also the reason why our model's performance in the ablation study is lowest when removing the semantic information (i.e. BERT encoding).

- **Semantic level**: Some other failed cases share the common features that the reply does not refer to the specific term mentioned in the quotation, but gives out a more general rebuttal, e.g. [quotation] *If the president is either killed or resigns, the vice president is a horrible choice to take over the office.* [reply] *Seriously, stop this hyperbole.* Our model cannot effectively distinguish the interactivity between them, as the reply is short and has entirely no overlapping with the quotation.
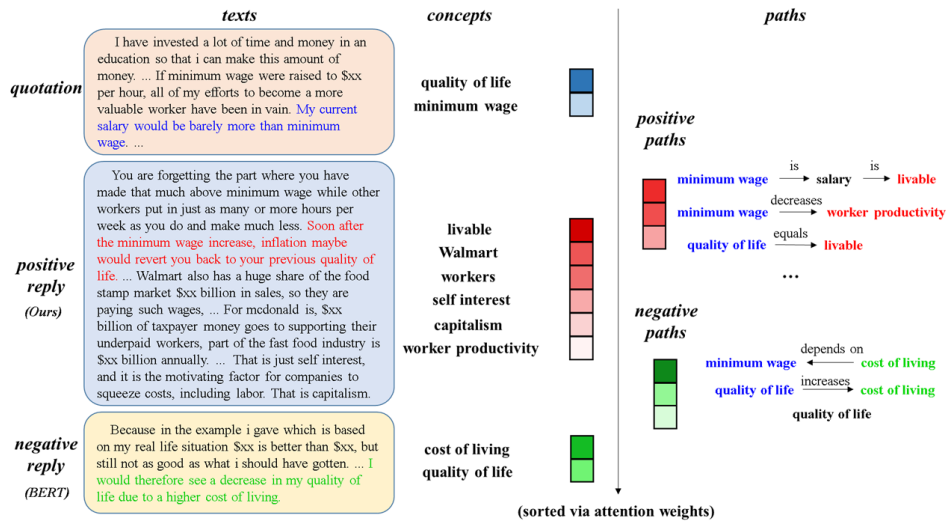
2316

Figure 4: Case study of an instance from the test set. Our model successfully chooses the correct reply argument from the given candidates while the BERT model gives the wrong answer. Note that the concepts and the paths are all sorted from top to bottom in terms of hierarchical attention weights.

5.3 Case study

A case study is as shown in Figure 4, where the negative reply is selected by BERT baseline. It shows that although the quotation and the negative reply share a common concept, *quality of life*, our model successfully figures out the interactive reply argument through the reasoning paths between the concepts from the two sides. All the concepts and the paths in the figure are arranged from top to bottom according to their respective attention weights. We can find that upper paths are actually highly related to the reasoning process of humans, and the irrelevant concepts such as worker productivity will automatically diminish by our hierarchical attention alignment.

6 Related Work

Dialogical argumentation mining As mentioned in Introduction(§1), our work mainly focus on dialogical argumentation mining. Among recent researches in this aspect, El Baff et al. (2020) compare content- and style-oriented classifiers on editorials to explore the effect of the writing style of editorials to the audience of different parties; Ji et al. (2019) propose the task of identifying interactive argument pairs in online debate forum such as ChangeMyView (CMV). Cheng et al. (2020) collects the text data from peer review and rebuttal process to mine the argumentative relationship entailed in such discussion; Khatib et al. (2020) constructs a monological argumentation graph by

extracting knowledge from *Debatepedia.org* and use human annotation to further improve the quality of their knowledge graph. Our work obtains inspiration from the construction of Al-Khatib's knowledge graph, but adapting their method to the dialogical debating forum settings, and removing the human annotation stage to obtain an automatically generated knowledge graph.

Leveraging external knowledge in NLU Our work also lies in the general context of using external knowledge to encode sentences and paragraphs. Yang and Mitchell (2017) are among the first researches that retrieve the related entities in the external knowledge base and merge them into an LSTM encoder. Afterward, Weissenborn et al. (2017), Mihaylov and Frank (2018) and Zhang et al. (2020) mainly follows the main idea of the work to incorporate external word-level lexical knowledge to enhance the sentence embedding. Moreover, Lin et al. (2019) propose a knowledge-aware network(KagNet) that utilizes that graph knowledge from *ConceptNet* to answer the commonsense questions. Compared with these methods, our work utilizes conceptual information from dialogical argumentation lexicons and conducts a reasoning process resembling human beings, which is then encoded by a path transformer, and finally aligned with the semantic information through a hierarchical attention mechanism.

7 Conclusion and Future Work

We propose a framework that imitates human's reasoning process in debating. Practically, we first construct a dialogical argumentation knowledge graph from the online debating forum Change-MyView, by using an automatic OpenIE toolkit and conducting concept grounding with lexical resources and Wikipedia API. Then we use a path-based graph model to encode the concepts and the reasoning path between concepts from two sides of a debate and align the conceptual information with the semantic information obtained implicitly by pre-trained language model BERT. Experiments on interactive argument pair identification task show that our model can leverage the external knowledge in both effective and transparent way.

Acknowledgements

References

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. Exploiting personal characteristics of debaters for predicting persuasiveness. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7067–7072, Online. Association for Computational Linguistics.

Christa SC Asterhan and Baruch B Schwarz. 2007. The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of educational psychology*, 99(3):626.

Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. 2020a. From arguments to key points: Towards automatic argument summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4029–4039, Online. Association for Computational Linguistics.

Roy Bar-Haim, Yoav Kantor, Lilach Eden, Roni Friedman, Dan Lahav, and Noam Slonim. 2020b. Quantitative argument summarization and beyond: Cross-domain key point analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 39–49, Online. Association for Computational Linguistics.

Rich Caruana, Steve Lawrence, and C. Lee Giles. 2000. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000, Denver, CO, USA*, pages 402–408. MIT Press.

Liying Cheng, Lidong Bing, Qian Yu, Wei Lu, and Luo Si. 2020. 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 (EMNLP)*, pages 7000–7011.

Yoonjung Choi and Janyce Wiebe. 2014. +/-EffectWordNet: Sense-level lexicon acquisition for opinion inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1181–1191, Doha, Qatar. Association for Computational Linguistics.

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

Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2020. Analyzing the Persuasive Effect of Style in News Editorial Argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3154–3160, Online. Association for Computational Linguistics.

Paolo Ferragina and Ugo Scaiella. 2010. TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, pages 1625–1628. ACM.

Lu Ji, Zhongyu Wei, Jing Li, Qi Zhang, and Xuanjing Huang. 2019. Discrete argument representation learning for interactive argument pair identification. *arXiv preprint arXiv:1911.01621*.

Jun Seok Kang, Song Feng, Leman Akoglu, and Yejin Choi. 2014. ConnotationWordNet: Learning connotation over the Word+Sense network. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1544–1554, Baltimore, Maryland. Association for Computational Linguistics.

Khalid Al Khatib, Yufang Hou, Henning Wachsmuth, Charles Jochim, Francesca Bonin, and Benno Stein. 2020. End-to-end argumentation knowledge graph construction. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 7367–7374. AAAI Press.

Jialu Li, Esin Durmus, and Claire Cardie. 2020. Exploring the role of argument structure in online debate persuasion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8905–8912, Online. Association for Computational Linguistics.

Bill Yuchen Lin, Xinyue Chen, Jamin Chen, and Xiang Ren. 2019. KagNet: Knowledge-aware graph networks for commonsense reasoning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2829–2839, Hong Kong, China. Association for Computational Linguistics.

Todor Mihaylov and Anette Frank. 2018. Knowledgeable reader: Enhancing cloze-style reading comprehension with external commonsense knowledge. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 821–832, Melbourne, Australia. Association for Computational Linguistics.

Gaku Morio, Hiroaki Ozaki, Terufumi Morishita, Yuta Koreeda, and Kohsuke Yanai. 2020. Towards better non-tree argument mining: Proposition-level biaffine parsing with task-specific parameterization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3259–3266, Online. Association for Computational Linguistics.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 311–321, Berlin, Germany. Association for Computational Linguistics.

Jason Tyler Rolfe. 2017. Discrete variational autoencoders. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4149–4158, Minneapolis, Minnesota. Association for Computational Linguistics.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Zhongyu Wei, Yang Liu, and Yi Li. 2016. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 195–200, Berlin, Germany. Association for Computational Linguistics.

Dirk Weissenborn, Tomáš Kočiskỳ, and Chris Dyer. 2017. Dynamic integration of background knowledge in neural nlu systems. *arXiv preprint arXiv:1706.02596*.

Bishan Yang and Tom Mitchell. 2017. Leveraging knowledge bases in LSTMs for improving machine reading. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1436–1446, Vancouver, Canada. Association for Computational Linguistics.

Jian Yuan, Zhongyu Wei, Yixu Gao, Wei Chen, Jun Song, Donghua Zhao, Heng Wang, Zhen Hu, Shaokun Zou, Donghai Li, et al. 2021. Overview of smp-cail2020-argmine: The interactive argument-pair extraction in judgement document challenge. *Data Intelligence*, pages 1–20.

Bowen Zhang, Xiaofei Xu, Min Yang, Xiaojun Chen, and Yunming Ye. 2018. Cross-domain sentiment classification by capsule network with semantic rules. *IEEE Access*, 6:58284–58294.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197, Online. Association for Computational Linguistics.