

Enhancing Event Causality Identification with Event Causal Label and Event Pair Interaction Graph

Ruili Pu¹, Yang Li², Suge Wang^{1,3*}, Deyu Li^{1,3}, Jianxing Zheng¹, Jian Liao¹

1. School of Computer and Information Technology, Shanxi University, Taiyuan, China

2. School of Finance, Shanxi University of Finance and Economics, Taiyuan, China

3. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China

{pu_ruili, liyangprimrose}@163.com, {wsg, lidy, jxzheng, liaoj}@sxu.edu.cn

Abstract

Most existing event causality identification (ECI) methods rarely consider the event causal label information and the interaction information between event pairs. In this paper, we propose a framework to enrich the representation of event pairs by introducing the event causal label information and the event pair interaction information. In particular, 1) we design an event-causal-label-aware module to model the event causal label information, in which we design the event causal label prediction task as an auxiliary task of ECI, aiming to predict which events are involved in the causal relationship (we call them causality-related events) by mining the dependencies between events. 2) We further design an event pair interaction graph module to model the interaction information between event pairs, in which we construct the interaction graph with event pairs as nodes and leverage graph attention mechanism to model the degree of dependency between event pairs. The experimental results show that our approach outperforms previous state-of-the-art methods on two benchmark datasets EventStoryLine and Causal-TimeBank.

1 Introduction

Event causality identification (ECI) aims to identify the causal relationship between pairs of events in the text. As shown in Figure 1(a), giving text and events as inputs, the ECI model needs to identify three causal relationships $\langle e7, \textit{cause}, e5 \rangle$, $\langle e5, \textit{cause}, e2 \rangle$, and $\langle e7, \textit{cause}, e2 \rangle$. As a semantic relationship, causality is important for semantic understanding and discourse analysis. Moreover, causal knowledge identified from a text can be useful for many natural language processing tasks (Fei et al., 2020a,b; Wang et al., 2020; Zhou et al., 2021; Dalal et al., 2021).

Various approaches have been proposed for ECI, from the early feature-based methods (Mirza, 2014;

*Corresponding author.

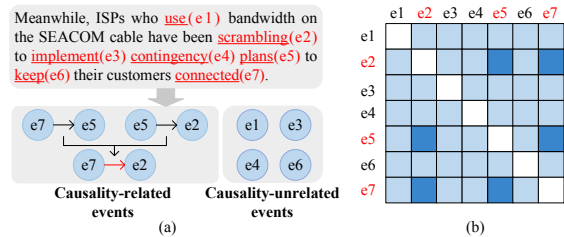


Figure 1: (a) An example of leveraging the event causal label information and the interaction information between event pairs for ECI. (b) The event pairs composed of seven events in (a), where the dark area denotes the event pairs composed of causality-related events.

Caselli and Vossen, 2017; Gao et al., 2019) to the current neural-network-based methods (Liu et al., 2020; Zuo et al., 2020, 2021a,b; Cao et al., 2021; Phu and Nguyen, 2021). The existing methods have achieved impressive performance. However, as far as we know, they usually improve the performance of ECI models by introducing commonsense knowledge or generating additional training data through data augmentation, with less focus on the inherent characteristics of the data. Specifically, 1) **Minimal use of event causal label information.** In the ECI task, a sentence usually contains multiple events. However, not all events are involved in causality. For instance, the sentence shown in Figure 1(a) has seven events, among which $e2$, $e5$, and $e7$ are causality-related events, whereas $e1$, $e3$, $e4$, and $e6$ are causality-unrelated events. In general, there is a strong semantic correlation between events that have a causal relationship. If the model can first identify $e2$, $e5$, and $e7$ as causality-related events according to the dependencies between events and contextual semantics, it can limit the identification scope of causality to some extent, as shown in Figure 1(b), thereby reducing the interference of unrelated events. Furthermore, by performing statistical analysis on two benchmark datasets, EventStoryLine and Causal-TimeBank, we find that 55.13% and 38.63% of the sentences with events in the two

datasets contained four or more events, where the average number of the different candidate event pairs is 13.12 and 10.64, and the average number of the event pairs with causal relationship is 1.99 and 0.30, respectively. If we can first predict which events are causality-related events according to the dependencies between them, the average number of candidate event pairs composed of causality-related events is 3.68 and 0.37, respectively, thereby narrowing the range of causality from 13.12 and 10.64 to 3.68 and 0.37, respectively. Therefore, it is necessary to introduce the event causal label information to assist the ECI tasks. 2) **Little focus on the interaction between event pairs.** The existing method (Phu and Nguyen, 2021) uses events as nodes and leverages toolkits to obtain information (such as dependency parsing tree and coreference) to construct event graph. However, they are modeled with events as nodes, cannot directly learn the interaction between event pairs. As shown in Figure 1(a), e_2 and e_7 are far from each other in the text. Thus, directly identifying the causal relationship between them is difficult for the model. If $\langle e_7, \text{cause}, e_5 \rangle$ and $\langle e_5, \text{cause}, e_2 \rangle$ are known, it is easier to infer the causal relationship between e_2 and e_7 according to the transitivity of the causal relationship (Hall, 2000). Therefore, it is essential to introduce the interaction information between event pairs to learn their dependencies.

To address the above limitations, we propose a framework called ECLP, which introduces Event Causal Label information and Event Pair interaction information to enrich the representation of event pairs. In particular, 1) we design an event-causal-label-aware module to model the event causal label information, in which we introduce the event causal label prediction task as an auxiliary task of ECI to mine the dependencies between events. 2) We further design an event pair interaction graph module to model the interaction information between event pairs. In particular, we construct the interaction graph with event pairs as nodes and adopt graph attention mechanism to model the degree of dependency between event pairs. The experimental results on two benchmark datasets show that our overall framework outperforms previous state-of-the-art methods.

2 Methodology

In ECI, the goal is to predict whether a causal relationship exists in each event pair $(e_i, e_j) (i \neq j)$

by giving text $S = (w_1, w_2, \dots, w_n)$ and events set $E = (e_1, e_2, \dots, e_m)$ contained in S as inputs, where w_i denotes the i -th word in S , and e_i denotes the i -th event in E . The overall framework of the proposed method is shown in Figure 2, which mainly includes four modules, we illustrate each component in detail in the following.

2.1 Encoding Layer

Given an input sentence S , we adopt a pre-trained language model BERT (Devlin et al., 2019) as the sentence encoder to extract hidden contextual representation $\mathbf{H}_S = (\mathbf{h}_{w_1}, \mathbf{h}_{w_2}, \dots, \mathbf{h}_{w_n})$, where \mathbf{h}_{w_i} denotes the hidden representation of the i -th token in S . Then, the events representation $\mathbf{H}_E = (\mathbf{h}_{e_1}, \mathbf{h}_{e_2}, \dots, \mathbf{h}_{e_m})$ can be obtained by summing the token representation contained in the event.

2.2 Event-Causal-Label-Aware Module

A sentence usually contains multiple events, however, not all events are causality-related events. We design an event-causal-label-aware module to mine the dependencies between events and help the model to pay attention to the extraction of the causal relationship from the event pairs composed of causality-related events. In general, there is a strong semantic correlation between events with causal relationships, so we first adopt the Transformer mechanism (Vaswani et al., 2017) to capture the dependencies between events, and get the updated event representation via Eq. (1).

$$\mathbf{H}_E^L = \text{Transformer}(\mathbf{H}_E) = (\mathbf{h}_{e_1}^L, \mathbf{h}_{e_2}^L, \dots, \mathbf{h}_{e_m}^L). \quad (1)$$

Then, we construct a binary classification to predict the probability $p_{e_i}^L$ of each event e_i in E as a causality-related event via Eq. (2).

$$p_{e_i}^L = \text{softmax}(\mathbf{W}_{e_i}^L \mathbf{H}_{e_i}^L + \mathbf{b}_{e_i}^L), \quad (2)$$

where $\mathbf{W}_{e_i}^L$ and $\mathbf{b}_{e_i}^L$ are learnable parameters.

To incorporate the event causal label into the event pair, we introduce a learnable label vector set $\mathbf{L} = \{l_{ij}\} (l_{ij} \in \{\mathbf{cr}, \mathbf{cu}\})$, where \mathbf{cr} denotes the event pairs composed of causality-related events, \mathbf{cu} denotes the event pairs that have at least one causality-unrelated event. In particular, the embedding vector \mathbf{L} is randomly initialized via sampling from a uniform distribution and is learned together with the model training process. For the event pair (e_i, e_j) , if e_i and e_j are both causality-related events (i.e. $p_{e_i}^L \geq 0.5$ and $p_{e_j}^L \geq 0.5$), the label of (e_i, e_j) is $l_{ij} = \mathbf{cr}$, else $l_{ij} = \mathbf{cu}$.

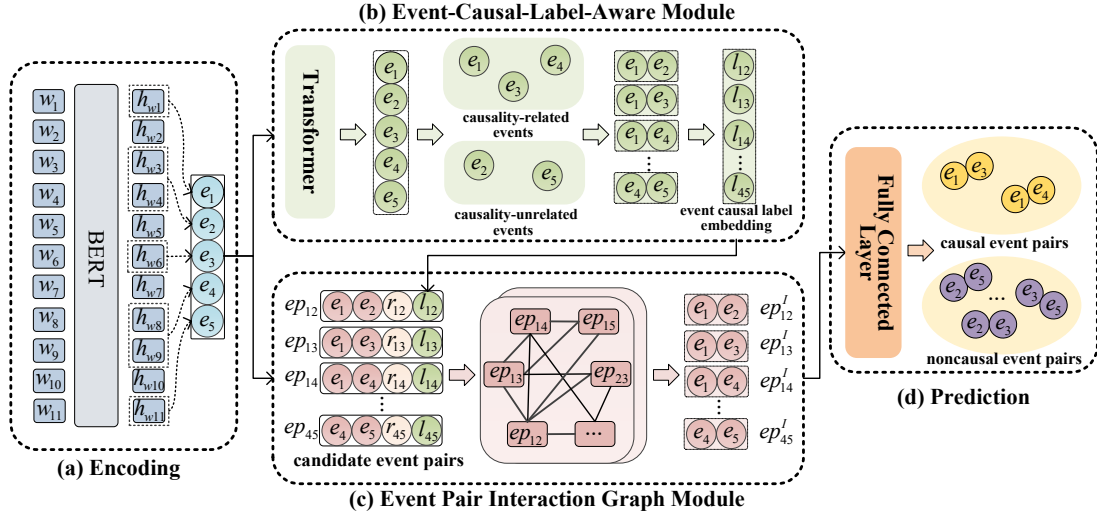


Figure 2: The framework of our proposed method.

2.3 Event Pair Interaction Graph Module

To capture the interaction information between event pairs, we construct the interaction graph with event pairs as nodes and adopt the graph attention mechanism to adaptively fuse the information of neighbor nodes.

• Event Pair Interaction Graph Construction.

Given the event set E , the events are combined in pairs to form the set of candidate causal event pairs $EP = \{ep_{ij}\} (0 < i, j \leq |E|, i \neq j)$ as nodes. In EP , the $ep_{ij} = [h_{e_i}; h_{e_j}; r_{ij}; l_{ij}]$, where h_{e_i} and $h_{e_j} \in H_E$ and r_{ij} denotes the relative position embedding between event e_i and event e_j in the text.

For the edges of the interaction graph, consider that event pairs in the same row or column are strongly associated with the current event pair (Ding et al., 2020). As shown in Figure 1(b), the relationship of event pair $\langle e_7, e_2 \rangle$ can be transmitted through the same row event pair $\langle e_7, e_5 \rangle$ and the same column event pair $\langle e_5, e_2 \rangle$. Thus, we connect the edges between the event pairs in the same row or column and add self-loop edges to fuse the information of their nodes.

• **Event Pair Interaction Graph Update.** Considering that different neighbor nodes have different importance for each event pair, we leverage Graph Attention Networks (GAT) (Veličković et al., 2018) to model the degree of dependency between event pairs. GAT propagates information among nodes by stacking multiple graph attention layers. At the t -th graph attention layer, the representation

of each node can be updated via Eq. (3).

$$ep_{ij}^t = \text{ReLU}\left(\sum_{uv \in \mathcal{N}(ij)} \alpha_{ij,uv}^t \mathbf{W}^t ep_{uv}^{t-1} + \mathbf{b}^t\right), \quad (3)$$

where $\mathcal{N}(ij)$ denotes the directly neighboring nodes of ep_{ij} ; the attention weight $\alpha_{ij,uv}^t$ is learned via Eq. (4), which reflects the strength of aggregation level between the nodes ep_{ij}^{t-1} and ep_{uv}^{t-1} ; \mathbf{W}^t , \mathbf{W}_{ij}^t , \mathbf{W}_{uv}^t , \mathbf{w}^t , and \mathbf{b}^t are learnable parameters.

$$\begin{aligned} ep_{ij,uv}^t &= \mathbf{w}^t \tanh([\mathbf{W}_{ij}^t ep_{ij}^{t-1}; \mathbf{W}_{uv}^t ep_{uv}^{t-1}]), \\ \alpha_{ij,uv}^t &= \frac{\exp(\text{LeakyReLU}(ep_{ij,uv}^t))}{\sum_{st \in \mathcal{N}(ij)} \exp(\text{LeakyReLU}(ep_{ij,st}^t))}, \end{aligned} \quad (4)$$

Therefore, the updated nodes representation $EP^I = \{ep_{ij}^I\}$ can be obtained by stacking T layers to model the inter-node relationships.

2.4 Prediction and Training

The predicted probability p_{ij} of the event pair (e_i, e_j) as a causal event pair can be obtained by performing a binary classification with ep_{ij}^I as input via Eq. (5).

$$p_{ij} = \text{softmax}(\mathbf{W}_{ij}^I ep_{ij}^I + \mathbf{b}_{ij}^I), \quad (5)$$

where \mathbf{W}_{ij}^I and \mathbf{b}_{ij}^I are learnable parameters.

For training, we utilize the cross-entropy function to supervise the causal event pair prediction via Eq. (6).

$$\mathcal{L}_{ep} = - \sum_{s \in D} \sum_{\substack{e_i, e_j \in E_s \\ e_i \neq e_j}} \mathbf{y}_{ij} \log(\mathbf{p}_{ij}), \quad (6)$$

where D denotes the training set; s denotes the sentence in D ; E_s denotes the events set in s ; \mathbf{y}_{ij} denotes the ground truth label of event pair (e_i, e_j) .

We also add an auxiliary supervision for the event causal label prediction task via Eq. (7).

$$\mathcal{L}_e = - \sum_{s \in D} \sum_{e_i \in E_s} \mathbf{y}_{e_i} \log(\mathbf{p}_{e_i}^L), \quad (7)$$

where $\mathbf{p}_{e_i}^L$ denotes the predicted probability of event e_i as a causality-related event; \mathbf{y}_{e_i} denotes the ground truth label of event e_i , which can be automatically obtained according to the labels of event pairs without any manual labeling.

The final loss function is a weighted sum of the aforementioned terms $\mathcal{L} = \mathcal{L}_{ep} + \lambda_e \mathcal{L}_e$, where $\lambda_e \in (0, 1)$.

3 Experiments

3.1 Datasets and Evaluation Metrics

We verify the effectiveness of the proposed model on two benchmark datasets, EventStoryLine (Caselli and Vossen, 2017) and Causal-TimeBank (Mirza, 2014), respectively. Following previous works (Cao et al., 2021; Zuo et al., 2021a), we perform the 5-fold and 10-fold cross-validation on the two datasets, respectively. In addition, we adopt Precision (P), Recall (R), and F1-score (F1) as evaluation metrics.

3.2 Parameter Settings

We use HuggingFace’s Transformers¹ library to implement the uncased BERT base model. The Adam algorithm (Kingma and Ba, 2015) is used as an optimizer, the learning rate is initialized to 2e-5, the batch size is set to 5, the GAT layers is set to 2, the dropout of GAT is set to 0.3, the dimensions of the event causal label embedding and relative position embedding are set to 80 and 40, respectively, and the weight λ_e is set to 0.2. Moreover, since the positive samples in the dataset are sparse, we adopt a negative sampling rate of 0.5 for training.

¹<https://github.com/huggingface/transformers>

3.3 Baselines

We compare our model with the following baseline methods. **Feature-based methods:** 1) **DD** (Mirza and Tonelli, 2014), a data-driven method. 2) **VR-C** (Mirza, 2014), a model with data filtering and causal signal enhancement. 3) **OP** (Caselli and Vossen, 2017), a dummy model to assign causality for ECI. 4) **Seq** (Choubey and Huang, 2017), a sequence model for ECI. 5) **ILP** (Gao et al., 2019), a document-level ECI model. **Neural-network-based methods:** 1) **KMMG** (Liu et al., 2020), which proposes a knowledge-aware reasoner and a mention masking reasoner for ECI. 2) **KnowDis** (Zuo et al., 2020), a distantly supervised method for ECI. 3) **LSIN** (Cao et al., 2021), a method that utilize the structural commonsense knowledge for ECI. 4) **LearnDA** (Zuo et al., 2021a), a learnable knowledge-guided data augmentation method for ECI. 5) **CauSeRL** (Zuo et al., 2021b), a self-supervised representation learning enhanced ECI method. 6) **RichGCN** (Phu and Nguyen, 2021), a graph convolutional network with rich information.

Model	EventStoryLine			Causal-TimeBank		
	P	R	F1	P	R	F1
DD	-	-	-	67.3	22.6	33.9
VR-C	-	-	-	69.0	31.5	43.2
OP	22.5	22.5	36.6	-	-	-
Seq	32.7	44.9	37.8	-	-	-
ILP	37.4	55.8	44.7	-	-	-
KMMG	41.9	62.5	50.1	36.6	55.6	44.1
KnowDis	39.7	66.5	49.7	42.3	60.5	49.8
CauSeRL	41.9	69.0	52.1	43.6	68.1	53.2
LSIN	47.9	58.1	52.5	51.5	56.2	52.9
LearnDA	42.2	69.8	52.6	41.9	68.0	51.9
RichGCN	49.2	63.0	55.2	39.7	56.5	46.7
ECLEP	49.3	68.1	57.1	50.6	63.4	56.3

Table 1: Experimental Results on the EventStoryLine and Causal-TimeBank Datasets.

3.4 Overall Results

The experimental results are shown in Table 1, we can observe that our proposed method ECLEP outperforms all the baselines on the two datasets. In particular, on the EventStoryLine, compared with the current best method RichGCN, our method ECLEP achieves 1.9% improvement in the F1-score; on the Causal-TimeBank, compared with the current best method CauSeRL, our method ECLEP achieves 3.1% improvement in the F1-score. This

finding indicates that our proposed method is effective for the ECI task. In addition, we observe that baseline methods of ECI often require external knowledge resources or toolkits to improve the performance. Our approach achieves the best performance by mining the inherent characteristics of the data.

3.5 Ablation Study

This section analyzes the contribution of each part in our model through ablation experiments, as shown in Table 2. In particular, we examine the following ablated models: 1) -ECL and -EPI denote the removal of the event causal label information and the event pair interaction information, respectively. We note that removing any of them degrades the performance for ECI, indicating that the two information we introduced are effective. 2) $-\mathcal{L}_e$ denote the removal of auxiliary supervision. We note that removing it degrades overall performance, because it can help the model more sufficiently learn the representation of the label vector. 3) -pos denote the removal of relative position embedding. The experimental results show that relative position information is helpful for ECI, because the probability of causal relationship between events that are closer is higher than those that are farther apart.

Model	P	R	F1
ECLEP	49.3	68.1	57.1
-ECL	48.1	65.7	55.5
-EPI	48.9	63.2	55.2
$-\mathcal{L}_e$	49.4	64.2	55.9
-pos	50.3	64.5	56.5

Table 2: Ablation Results on the EventStoryLine dataset.

3.6 Visualization Analysis

We visualize the distribution of each module to explore the effectiveness of our model further. The following can be observed from Figure 3: 1) The event causal label information and the event pair interaction information focus on different aspects of features to identify the causal relationships and they share complementary effects. This finding also provides an explanation on the good performance of our full model ECLEP. 2) The event causal label can limit the identification scope of causality to some extent, and help the model to pay attention to the extraction of the causal relationship from the

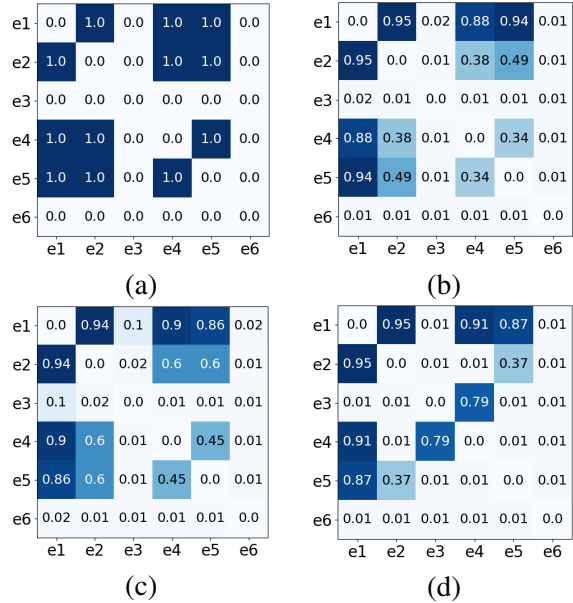


Figure 3: (a) Distribution of the label L in the event-causal-label-aware module, where the shaded areas denote the event pairs composed of causality-related events. (b) Distribution of the full model ECLEP. (c) Distribution that only using the event causal label information. (d) Distribution that only using the event pair interaction information. In (b), (c), and (d), the shaded area denotes the probability of each event pair is predicted as a causal event pair. In this example, the ground truth labels are $(e1, e2)$, $(e1, e4)$ and $(e1, e5)$.

event pairs composed of causality-related events.

4 Conclusion

In this paper, we propose a framework to enrich the representation of event pairs by introducing event causal label information and event pair interaction information. The experimental results on two widely used datasets indicate that our approach is effective for the ECI task. In the future, we aim to mine other potential causal features for this task and apply our model to other types of relation extraction tasks, such as temporal relation extraction.

Limitations

In this paper, we only focus on whether or not there is a causal relationship between the given events, does not discriminate the specific cause/effect event. In addition, we only conduct research on sentence-level ECI, whereas document-level ECI often present more challenges. These are the focus of our future research.

Acknowledgements

The authors would like to thank all anonymous reviewers for their valuable comments and suggestions which have significantly improved the quality and presentation of this paper. The works described in this paper are supported by the National Natural Science Foundation of China (62106130, 62076158, 62072294, 62272286), Natural Science Foundation of Shanxi Province, China (20210302124084), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi, China (2021L284), and CCF-Zhipu AI Large Model Foundation of China (CCF-Zhipu202310).

References

- Pengfei Cao, Xinyu Zuo, Yubo Chen, Kang Liu, Jun Zhao, Yuguang Chen, and Weihua Peng. 2021. [Knowledge-enriched event causality identification via latent structure induction networks](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4862–4872, Online. Association for Computational Linguistics.
- Tommaso Caselli and Piek Vossen. 2017. The event storyline corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86.
- Prafulla Kumar Choubey and Ruihong Huang. 2017. [A sequential model for classifying temporal relations between intra-sentence events](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1796–1802, Copenhagen, Denmark. Association for Computational Linguistics.
- Dhairya Dalal, Mihael Arcan, and Paul Buitelaar. 2021. Enhancing multiple-choice question answering with causal knowledge. In *Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 70–80.
- 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.
- Zixiang Ding, Rui Xia, and Jianfei Yu. 2020. [ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3161–3170, Online. Association for Computational Linguistics.
- Hao Fei, Yafeng Ren, and Donghong Ji. 2020a. Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction. *Information Processing & Management*, 57(6):102311.
- Hao Fei, Yafeng Ren, and Donghong Ji. 2020b. A tree-based neural network model for biomedical event trigger detection. *Information Sciences*, 512:175–185.
- Lei Gao, Prafulla Kumar Choubey, and Ruihong Huang. 2019. [Modeling document-level causal structures for event causal relation identification](#). 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 1808–1817, Minneapolis, Minnesota. Association for Computational Linguistics.
- Ned Hall. 2000. Causation and the price of transitivity. *The Journal of Philosophy*, 97(4):198–222.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *the 6th International Conference on Learning Representations*.
- Jian Liu, Yubo Chen, and Jun Zhao. 2020. Knowledge enhanced event causality identification with mention masking generalizations. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 3608–3614.
- Paramita Mirza. 2014. [Extracting temporal and causal relations between events](#). In *Proceedings of the ACL 2014 Student Research Workshop*, pages 10–17, Baltimore, Maryland, USA. Association for Computational Linguistics.
- Paramita Mirza and Sara Tonelli. 2014. [An analysis of causality between events and its relation to temporal information](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2097–2106, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Minh Tran Phu and Thien Huu Nguyen. 2021. [Graph convolutional networks for event causality identification with rich document-level structures](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3480–3490, Online. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30:5998–6008.

- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *the 6th International Conference on Learning Representations*.
- Jin Wang, Chunbin Lin, Mingda Li, and Carlo Zaniolo. 2020. Boosting approximate dictionary-based entity extraction with synonyms. *Information Sciences*, 530:1–21.
- Guangyou Zhou, Zhiwen Xie, Zongfu Yu, and Jimmy Xiangji Huang. 2021. Dfm: A parameter-shared deep fused model for knowledge base question answering. *Information Sciences*, 547:103–118.
- Xinyu Zuo, Pengfei Cao, Yubo Chen, Kang Liu, Jun Zhao, Weihua Peng, and Yuguang Chen. 2021a. [LearnDA: Learnable knowledge-guided data augmentation for event causality identification](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3558–3571, Online. Association for Computational Linguistics.
- Xinyu Zuo, Pengfei Cao, Yubo Chen, Kang Liu, Jun Zhao, Weihua Peng, and Yuguang Chen. 2021b. [Improving event causality identification via self-supervised representation learning on external causal statement](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2162–2172, Online. Association for Computational Linguistics.
- Xinyu Zuo, Yubo Chen, Kang Liu, and Jun Zhao. 2020. [KnowDis: Knowledge enhanced data augmentation for event causality detection via distant supervision](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1544–1550, Barcelona, Spain (Online). International Committee on Computational Linguistics.

ACL 2023 Responsible NLP Checklist

A For every submission:

A1. Did you describe the limitations of your work?
5

A2. Did you discuss any potential risks of your work?
There is no risk involved in our work.

A3. Do the abstract and introduction summarize the paper's main claims?
1

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

B Did you use or create scientific artifacts?

Left blank.

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

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

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

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

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

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

C Did you run computational experiments?

3

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

3

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

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

3

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

3

- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?

3

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

Left blank.

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

No response.

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

No response.

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?

No response.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

No response.

- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

No response.