A Joint Framework with Heterogeneous-Relation-Aware Graph and Multi-Channel Label Enhancing Strategy for Event Causality Extraction

Ruili Pu[1], **Yang Li**[2], **Jun Zhao**[3], **Suge Wang**[1,4*], **Deyu Li**[1,4], **Jian Liao**[1], **Jianxing Zheng**[1]

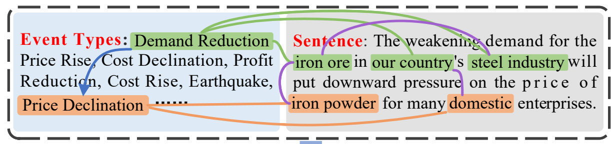
[1]School of Computer and Information Technology, Shanxi University, Taiyuan, China
[2]School of Finance, Shanxi University of Finance and Economics, Taiyuan, China
[3]National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China
[4]Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China
{pu_ruili, liyangprimrose}@163.com, jzhao@nlpr.ia.ac.cn, {wsg, lidy, jxzheng, liaoj}@sxu.edu.cn

Abstract

Event Causality Extraction (ECE) aims to extract the cause-effect event pairs with their structured event information from plain texts. As far as we know, the existing ECE methods mainly focus on the correlation between arguments, without explicitly modeling the causal relationship between events, and usually design two independent frameworks to extract cause events and effect events, respectively, which cannot effectively capture the dependency between the subtasks. Therefore, we propose a joint multi-label extraction framework for ECE to alleviate the above limitations. In particular, 1) we design a heterogeneous-relation-aware graph module to learn the potential relationships between events and arguments, in which we construct the heterogeneous graph by taking the predefined event types and all the words in the sentence as nodes, and modeling three relationships of "event-event", "event-argument" and "argument-argument" as edges. 2) We also design a multi-channel label enhancing module to better learn the distributed representation of each label in the multi-label extraction framework, and further enhance the interaction between the subtasks by considering the preliminary results of cause-effect type identification and event argument extraction. The experimental results on the benchmark dataset ECE-CCKS show that our approach outperforms previous state-of-the-art methods, and that our model also performs well on the complex samples with multiple cause-effect event pairs.

Introduction

Event Causality Extraction (ECE) is an important task in the area of Natural Language Processing (NLP). The ECE task can be divided into two subtasks: ***Cause-Effect Type Identification*** (***CET***) and ***Event Argument Extraction*** (***EAE***) (Cui et al. 2022). As shown in Figure 1, given the predefined event types and sentence as inputs, the ECE model needs to identify the cause-effect type <*Demand Reduction*, *Price Declination*> and extract their structured event information, specifically, the arguments of cause event *Demand Reduction* are *our country*$_{Region}$, *steel industry*$_{Industry}$, and *iron ore*$_{Product}$, the arguments of effect event *Price Declination* are *domestic*$_{Region}$ and *iron powder*$_{Product}$. It can benefit many NLP tasks, such as event forecasting (Hashimoto 2019), question answering (Shi et al. 2021), and others.



Causality	Event Type	Event Roles		
		Region	Industry	Product
Cause	Demand Reduction	our country	steel industry	iron ore
Effect	Price Declination	domestic	—	iron powder

Figure 1: An example of ECE.

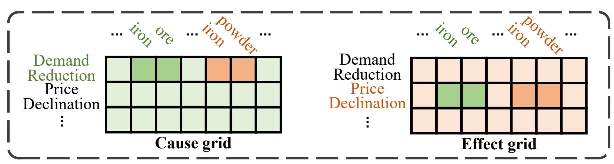


Figure 2: The dual grid tagging scheme of Cui et al. (2022)

In the ECE task, we note that the following three types of relationships are essential for understanding the task.

• ***Event-Event***: Modeling the relationships between events is helpful to capture the potential causality between them, which is the key to extract cause-effect event pairs. For example, the event "Earthquake" usually leads to some negative events, such as "Injury".

• ***Event-Argument***: Modeling the relationships between events and arguments can help to extract specific arguments based on each event type. Specifically, each event has its specific arguments, and as the event type changes, so does its arguments. As shown in Figure 1, for the event "Demand Reduction", the argument of the role "Product" is "iron ore", while for the event "Price Declination", the argument of the role "Product" is "iron powder". Therefore, it is helpful for the ECE model to better capture the interaction between events and arguments.

• ***Argument-Argument***: Arguments are the core components of events, and the relationships between events are

*Corresponding author.

also related to the relationships between their arguments. As shown in Figure 1, there is a strong semantic correlation between the argument "iron ore" in the cause event "Demand Reduction" and the argument "iron powder" in the effect event "Price Declination". In addition, there is also a strong semantic correlation between the arguments "iron ore" and "steel industry" in the same event "Demand Reduction". Therefore, modeling the relationships between arguments is help for the ECE model to learn the potential correlation between events and extract arguments of each event more reliably.

Furthermore, in the ECE task, the two subtasks CET and EAE are interdependent. As shown in Figure 1, if the ECE model predicts that <*Demand Reduction*, *Price Declination*> is cause-effect type, then it is more likely to extract cause arguments (such as *iron ore*) from the event *Demand Reduction* and effect arguments (such as *iron powder*) from the event *Price Declination*. Conversely, if event *Demand Reduction* is predicted to have cause arguments, and event *Price Declination* is predicted to have effect arguments, there is a higher probability that these two events are real events. Since they are in the same sentence and based on the data characteristics, the probability of a causal relationship between them is relatively high.

However, the existing methods have made limited use of the aforementioned features. Cui et al. (2022) designed a dual grid tagging scheme, as shown in Figure 2, they adopted two independent grids to extract the arguments of each cause event and effect event, respectively. Specifically, for the cause grid, in the row of event *Demand Reduction*, they not only predict *our country*, *steel industry*, and *iron ore* as the cause arguments according to the correlation between events and arguments, but also predict *domestic* and *iron powder* as the arguments of the corresponding effect event according to the correlation between arguments. The same applies to the effect grid. Then, the cause-effect event pairs can be obtained by matching the arguments co-occurring in both event grids simultaneously. Although they have achieved impressive performance, they mainly focus on the relationships of "event-argument" and "argument-argument", without explicitly modeling the relationship of "event-event", and they design two independent grids to model cause events and effect events separately, which cannot effectively capture the dependency between the two subtasks.

To address the above limitations, we propose a Joint multi-label Extraction Framework with Heterogeneous-relation-aware graph and Multi-channel label enhancing strategy (JEF-HM) for the ECE task. Specifically, 1) we design a heterogeneous-relation-aware graph module to learn the potential relationships between events and arguments, in which we construct the heterogeneous graph by taking the predefined event types and all the words in the sentence as nodes, and modeling three relationships of "event-event", "event-argument" and "argument-argument" as edges, then, we adopt the heterogeneous graph neural networks to learn the dependency between nodes. 2) We also design a multi-channel label enhancing module to utilize multiple channels to learn the distributed representations of each label, including event-event causality label and event-argument relation labels, and further enhance the interaction between the two subtasks EAE and CET by enhancing strategy.

In summary, our contributions are as follows:

- We propose a joint multi-label extraction framework to jointly extract the CET and EAE subtasks for ECE.
- To explore potential causal clues for ECE, we design two modules, namely, heterogeneous-relation-aware graph module and multi-channel label enhancing module.
- The experimental results on the benchmark dataset ECE-CCKS show that our approach outperforms previous state-of-the-art methods, and that our model also performs well on the complex samples with multiple cause-effect event pairs.

Related Work

In this section, we introduce event causality from the following two aspects:

Event Causality Identification

Given the rapid development of natural language processing, event causality has attracted increasing attention from researchers. Currently, most researchers mainly focus on the Event Causality Identification (ECI) task, which aims to identify the causal relationship between two given events in the text.

In earlier studies, feature-based methods were mainly used to solve the ECI tasks, including lexical and syntactic features (Gao, Choubey, and Huang 2019), causality cues (such as "cause" and "therefore") (Riaz and Girju 2014), event co-occurrence patterns (Hu and Walker 2017), temporal patterns (Ning et al. 2018), and others.

In recent years, many neural-network-based methods have been applied to the ECI task because of the widespread development of deep learning. Liu, Chen, and Zhao (2020), Cao et al. (2021), and Wu et al. (2023) introduced external commonsense knowledge to enrich the representations of events. Zuo et al. (2020) and Zuo et al. (2021) generated additional training data through data augmentation framework to alleviate data lacking problem. Shen et al. (2022) proposed a derivative prompt joint learning model, which utilized potential causal knowledge in the pre-trained language model to enhance the ECI model. Chen et al. (2022b) proposed a event relational graph transformer framework to capture the potential causal chains and alleviate the false positive and false negative issues for document-level ECI.

Event Causality Extraction

Although the ECI task has achieved impressive performance, it still suffers from limitations. For example, in the ECI task, each event is expressed by a word or phrase, there is no clear event type and event structure information, and events need to be provided in advance, which may lead to the loss of some crucial clues for understanding event causality and restricts its applicability. Therefore, Cui et al. (2022) further introduced the ECE task to overcome the above limitations of ECI task, and designed a dual grid tagging scheme for ECE, which adopts two independent grids to extract the

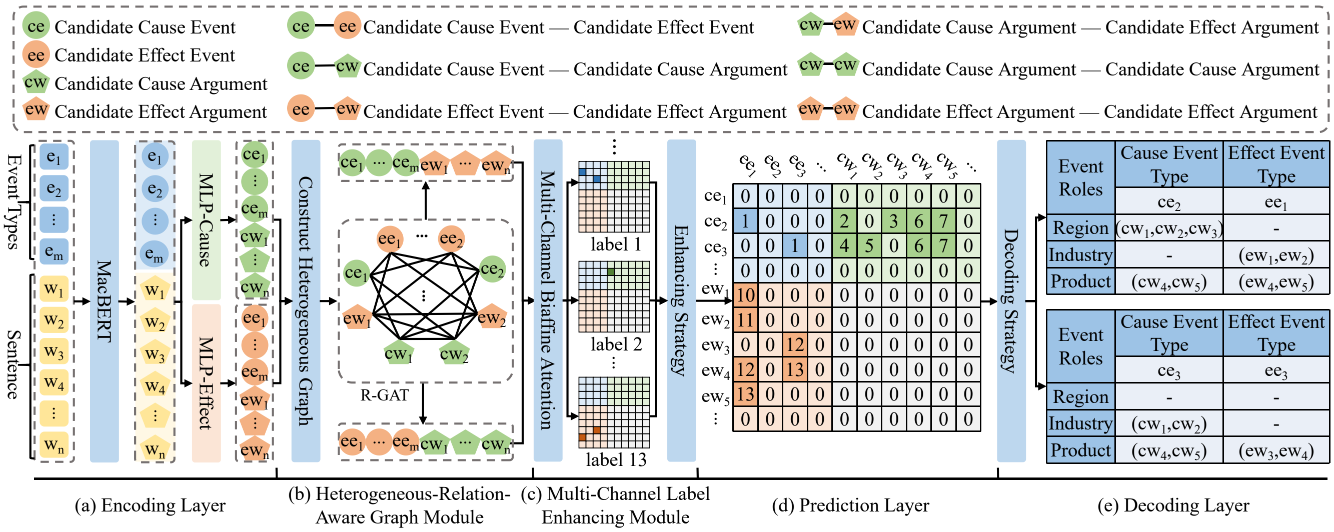


Figure 3: The framework of our proposed model JEF-HM.

arguments of each cause event and effect event, respectively, and obtains the cause-effect event pairs by matching the arguments in the two grids.

Despite their success, they do not explicitly model the causal relationship between events, and the two independent grids they designed are unable to effectively capture the dependency between two subtasks. Therefore, we propose a joint multi-label extraction framework with heterogeneous-relation-aware graph module and multi-channel label enhancing module to alleviate the above limitations.

Method

Problem Formulation

Given a predefined set $E = \{e_1, e_2, ..., e_m\}$ containing m event types and a sentence $S = \{w_1, w_2, ..., w_n\}$ containing n words as input, the goal of ECE task is to extract a set $O = \{(ce, ee, ca_{role}, ea_{role})_o\}_{o=1}^{o=|O|}$ from the input sequence, where ce and ee represent the cause event type and effect event type, respectively, ca_{role} and ea_{role} represent the sets of cause arguments and effect arguments, respectively, $role \in \{Region, Industry, Product\}$, $|O|$ represents the number of cause-effect event pairs in the sentence.

Label Definition and Grid Filling

We define 14 types of labels (as shown in Table 1) and construct a $(m+n)*(m+n)$ grid T (as shown in Figure 3(d)) for ECE. For the grid T, the filling process can be divided into the following three steps:

• **Cause-effect type filling**. The entry $T_{i,j}$ $(i, j \in [1, m])$ is used to fill cause-effect type. If there exists a causal relationship between event e_i (as cause event) and event e_j (as effect event), we set $T_{i,j} = 1$; otherwise, $T_{i,j} = 0$.

• **Cause event argument filling**. The entry $T_{i,j}$ $(i \in [1, m], j \in [m+1, m+n])$ is used to fill the argument of the cause event. If the $(j-m)$-th word $w_{(j-m)}$ is the argument boundary of the i-th cause event e_i, we set $T_{i,j} = ct$ $(ct \in [2, 7])$, otherwise, $T_{i,j} = 0$.

#	Labels	#	Labels
0	Null	7	Cause-Product-E
1	Cause-Effect	8	Effect-Region-S
2	Cause-Region-S	9	Effect-Region-E
3	Cause-Region-E	10	Effect-Industry-S
4	Cause-Industry-S	11	Effect-Industry-E
5	Cause-Industry-E	12	Effect-Product-S
6	Cause-Product-S	13	Effect-Product-E

Table 1: The label set $L = \{l_i\}(i \in [0, 13])$ of the cause-effect grid, where Cause-Effect denotes the event-event causality label; Cause-Role-S/E denotes the start/end position label of the argument under the corresponding role in the cause event; Effect-Role-S/E denotes the start/end position label of the argument under the corresponding role in the effect event.

• **Effect event argument filling**. The entry $T_{i,j}$ $(i \in [m+1, m+n], j \in [1, m])$ is used to fill the argument of the effect event. If the $(i-m)$-th word $w_{(i-m)}$ is the argument boundary of the j-th effect event e_j, we set $T_{i,j} = et$ $(et \in [8, 13])$, otherwise, $T_{i,j} = 0$.

JEF-HM Model

The overview of our proposed JEF-HM model is shown in Figure 3.

Encoding Given a predefined event type set E and sentence S, we concatenate them as input sequence via Eq. (1), and adopt a pre-trained language model MacBERT (Cui et al. 2020) as encoder to extract hidden contextual representations via Eq. (2). Then, we utilize two Multi-Layer-Perceptrons (MLP) to capture the cause-specific representations and effect-specific representations via Eqs. (3)-(4), respectively.

$$I = ([\text{CLS}], e_1, [\text{E}_1], ..., [\text{SEP}], w_1, ..., w_n, [\text{SEP}]) \quad (1)$$

$$H_I = \mathrm{MacBERT}(I) = [h_{e_1}, ..., h_{e_m}, h_{w_1}, ..., h_{w_n}] \quad (2)$$

$$H_C = \mathrm{MLP}_C(H_I) = [h_{ce_1}, ..., h_{ce_m}, h_{cw_1}, ..., h_{cw_n}] \quad (3)$$

$$H_E = \mathrm{MLP}_E(H_I) = [h_{ee_1}, ..., h_{ee_m}, h_{ew_1}, ..., h_{ew_n}] \quad (4)$$

where [$\mathrm{E_j}$] denotes the marker of the event type e_j; h_{ce_i}, h_{ee_i}, h_{cw_i}, and h_{ew_i} denote the representations of the candidate cause event, candidate effect event, candidate cause argument, and candidate effect argument, respectively.

Heterogeneous-Relation-Aware Graph Module To capture the potential relationships between events and arguments, we construct a heterogeneous-relation-aware graph and adopt the heterogeneous graph neural networks to adaptively fuse the information of neighbor nodes.

• **Heterogeneous-Relation-Aware Graph Construction.** We define the graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the node set $\mathcal{V}$ consists of four parts, including candidate cause event set $\{ce_i\}_{i=1}^m$, candidate effect event set $\{ee_i\}_{i=1}^m$, candidate cause argument set $\{cw_i\}_{i=1}^n$, and candidate effect argument set $\{ew_i\}_{i=1}^n$. Therefore, the embeddings of all nodes can be obtained via Eq. (5).

$$H_{\mathcal{V}}^{(0)} = \{h_{ce_i}\}_{i=1}^m \cup \{h_{ee_i}\}_{i=1}^m \cup \{h_{cw_i}\}_{i=1}^n \cup \{h_{ew_i}\}_{i=1}^n \quad (5)$$

For the edge set $\mathcal{E}$ of the graph, we consider the following three relationships:

1) Event-Event: Modeling the relationships between events is conducive to identify their potential causal relationship. Therefore, we construct edges for each candidate cause event ce_i and candidate effect event ee_i.

2) Event-Argument: Modeling the relationships between events and arguments can help to better capture the interaction between them. There are two kinds of Event-Argument edges in our graph. Specifically, for cause, we establish edges between each candidate cause event ce_i and candidate cause argument cw_i; for effect, we establish edges between each candidate effect event ee_i and candidate effect argument ew_i.

3) Argument-Argument: Modeling the relationships between arguments can help to mine the inherent correlation between events and extract arguments of each event more reliably. There are three kinds of Argument-Argument edges in our graph. Specifically, we establish edges between each candidate cause argument cw_i and candidate effect argument ew_i; furthermore, considering that the correlation between two words with semantic relation is relatively strong, we utilize the toolkit LTP (Che et al. 2021) to obtain syntactic dependency tree and establish edges between candidate cause arguments cw_i and cw_j which are connected to each other in dependency tree, the same applies to the candidate effect arguments.

Furthermore, each type of nodes have a self-loop edge, which can help each node to maintain its feature in the process of interaction.

• **Heterogeneous-Relation-Aware Graph Update.** Considering that different neighbor nodes and different types of edges have different importance for each node, we leverage Relational Graph Attention Network (R-GAT) (Wang et al. 2020a) to model the degree of dependency between nodes.

R-GAT contains Q attentional heads and K relational heads. Specifically, attentional heads update each node representation by using multi-head attention via Eqs. (6)-(7).

$$h_{att_i}^{(l+1)} = ||_{q=1}^{Q} \sum_{v_j \in \mathcal{N}(v_i)} \alpha_{i,j}^{lq} W_q^l h_{v_j}^l \quad (6)$$

$$\alpha_{i,j}^{lq} = attention(i, j) \quad (7)$$

where $\mathcal{N}(v_i)$ denotes the neighbor nodes of v_i; $\alpha_{i,j}^{lq}$ is a normalized attention coefficient, following Wang et al. (2020a), we adopt dot-product attention; W_q^l is a learnable parameter.

Relational heads update each node representation by considering different types of dependency relations via Eqs. (8)-(10).

$$h_{rel_i}^{(l+1)} = ||_{k=1}^{K} \sum_{v_j \in \mathcal{N}(v_i)} \beta_{i,j}^{lk} W_k^l h_{v_j}^l \quad (8)$$

$$\beta_{i,j}^{lk} = \frac{\exp(g_{i,j}^{lk})}{\sum_{u=1}^{\mathcal{N}(v_i)} \exp(g_{i,u}^{lk})} \quad (9)$$

$$g_{i,j}^{lk} = \sigma(\mathrm{ReLU}(r_{i,j} W_{k_1} + b_{k_1}) W_{k_2} + b_{k_2}) \quad (10)$$

where $r_{i,j}$ denotes the relation embedding between nodes v_i and v_j; W_k^l, W_{k_1}, W_{k_2}, b_{k_1}, and b_{k_2} are learnable parameters.

The representation of each node at layer $l+1$ can be computed via Eq. (11).

$$h_{v_i}^{(l+1)} = \mathrm{ReLU}(W_{l+1}[h_{att_i}^{(l+1)}; h_{rel_i}^{(l+1)}] + b_{l+1}) \quad (11)$$

where W_{l+1} and b_{l+1} are learnable parameters.

In order to preserve the contextual semantic information of each node, the final representations are a weighted sum of the original representations $H_{\mathcal{V}}^{(0)}$ and the last layer's representations $H_{\mathcal{V}}^{(l_n)}$ via Eq. (12).

$$\widetilde{H}_{\mathcal{V}} = W_{l_0} H_{\mathcal{V}}^{(0)} + W_{l_n} H_{\mathcal{V}}^{(l_n)} \quad (12)$$

where $H_{\mathcal{V}}^{(l_n)}$=$\{h_{v_i}^{(l_n)}\}_{i=1}^{2*(m+n)}$; W_{l_0} and W_{l_n} are learnable parameters.

Multi-Channel Label Enhancing Module To better learn the distributed representation of each label in Table 1, and further enhance the interaction between the two subtasks, inspired by Chen et al. (2022a), we design a multi-channel biaffine attention module and enhancing strategy, respectively.

• **Multi-Channel Biaffine Attention.** To construct attention matrix, the column representations $\widetilde{H}_C$ and the row representations $\widetilde{H}_E$ can be obtained via Eqs. (13)-(14).

$$\widetilde{H}_C = \{\widetilde{h}_{ce_i}\}_{i=1}^m \cup \{\widetilde{h}_{ew_i}\}_{i=1}^n \quad (13)$$

$$\widetilde{H}_E = \{\widetilde{h}_{ee_i}\}_{i=1}^m \cup \{\widetilde{h}_{cw_i}\}_{i=1}^n \quad (14)$$

where $\widetilde{h}_{ce_i}$, $\widetilde{h}_{ee_i}$, $\widetilde{h}_{cw_i}$ and $\widetilde{h}_{ew_i} \in \widetilde{H}_{\mathcal{V}}$.

Then, we adopt a biaffine attention mechanism (Dozat and Manning 2017) to generate the multi-channel label representations via Eqs. (15)-(16).

$$lr_{i,j,t} = \widetilde{h}_i^{c\mathrm{T}} U_1 \widetilde{h}_j^e + U_2[\widetilde{h}_i^c; \widetilde{h}_j^e] + b \quad (15)$$

$$lr_{i,j} = [lr_{i,j,1}; lr_{i,j,2}; ..., lr_{i,j,|L|}] \tag{16}$$

where $lr_{i,j,t}$ denotes the distributed representation of the t-th label l_t; $|L|$ denotes the number of label; $\widetilde{h}_i^c \in \widetilde{H}_C$; $\widetilde{h}_j^e \in \widetilde{H}_E$; U_1, U_2 and b are learnable parameters.

• **Enhancing Strategy.** To further enhance the interaction between the two subtasks, we design an effective enhancing strategy by utilizing the event-argument relationships to strengthen the event-event causal relationship, while also using the event-event causal relationship to further enhance the event-argument relationships. Specifically, we introduce lr_i^{cw} and lr_j^{ew} as the preliminary results of EAE when identifying the causal relationship between events, and introduce lr_i^{ce} and lr_j^{ee} as the preliminary results of the CET when conducting argument extraction via Eqs. (17)-(23).

For cause-effect type identification:

$$lr_i^{cw} = \text{max\_pooling}(lr_{i,m+1:m+n}) \tag{17}$$

$$lr_j^{ew} = \text{max\_pooling}(lr_{m+1:m+n,j}) \tag{18}$$

$$h_{i,j} = (\widetilde{h}_i^c; \widetilde{h}_j^e; lr_{i,j}; lr_i^{cw}; lr_j^{ew}) \tag{19}$$

where $i, j \in [1, m]$.

For cause event argument extraction:

$$lr_i^{ce} = \text{max\_pooling}(lr_{i,1:m}) \tag{20}$$

$$h_{i,j} = (\widetilde{h}_i^c; \widetilde{h}_j^e; lr_{i,j}; lr_i^{ce}) \tag{21}$$

where $i \in [1, m], j \in [m + 1, m + n]$.

For effect event argument extraction:

$$lr_j^{ee} = \text{max\_pooling}(lr_{1:m,j}) \tag{22}$$

$$h_{i,j} = (\widetilde{h}_i^c; \widetilde{h}_j^e; lr_{i,j}; lr_j^{ee}) \tag{23}$$

where $i \in [m + 1, m + n], j \in [1, m]$.

Prediction and Training After obtaining the representation $h_{i,j}$ of each entry in the grid, we conduct multi-label classification upon each entry via Eq. (24).

$$p_{i,j} = \sigma(W_p h_{i,j} + b_p) \tag{24}$$

where W_p and b_p are learnable parameters.

For training, we utilize the cross-entropy function as loss function via Eq. (25).

$$\begin{aligned} \mathcal{L}_p = -(&\sum_{i=1}^{m}\sum_{j=1}^{m+n}\sum_{t\in|L|} \mathrm{I}(y_{i,j} = l_t)\log(p_{i,j}|l_t) \\ &+ \sum_{i=m}^{m+n}\sum_{j=1}^{m}\sum_{t\in|L|} \mathrm{I}(y_{i,j} = l_t)\log(p_{i,j}|l_t)) \end{aligned} \tag{25}$$

where $y_{i,j}$ denotes the ground truth label of the entry $T_{i,j}$.

We also add an auxiliary supervision for the multi-channel label distributed representations via Eq. (26).

$$\begin{aligned} \mathcal{L}_{lr} = -(&\sum_{i=1}^{m}\sum_{j=1}^{m+n}\sum_{t\in|L|} \mathrm{I}(y_{i,j} = l_t)\log(\sigma(lr_{i,j})|l_t) \\ &+ \sum_{i=m}^{m+n}\sum_{j=1}^{m}\sum_{t\in|L|} \mathrm{I}(y_{i,j} = l_t)\log(\sigma(lr_{i,j})|l_t)) \end{aligned} \tag{26}$$

The final loss function is a weighted sum of the aforementioned terms via Eq. (27).

$$\mathcal{L} = \mathcal{L}_p + \beta\mathcal{L}_{lr} \tag{27}$$

where $\beta \in (0, 1)$.

Decoding Strategy The decoding strategy can be divided into the following two steps:

• **Cause-effect type decoding**. The predicted results of $\{p_{i,j}\}(i, j \in [1, m])$ are used to decode cause-effect types. As shown in Figure 3(d), we can obtain cause-effect types set $CE = \{(ce_2, ee_1), (ce_3, ee_3)\}$.

• **Cause/effect event argument decoding**. For each cause-effect type $(ce_i, ee_j) \in CE$, we adopt the nearest start-end match principle (Wei et al. 2020) to decode the cause arguments from $\{p_{ce_i,j}\}(j \in [m + 1, m + n])$ and the effect arguments from $\{p_{i,ee_j}\}(i \in [m + 1, m + n])$. As shown in Figure 3(d), for (ce_2, ee_1), we can obtain the arguments $\{(cw_1, cw_2, cw_3)_{Region}, (cw_4, cw_5)_{Product}\}$ for the cause event ce_2, and the arguments $\{(ew_1, ew_2)_{Industry}, (ew_4, ew_5)_{Product}\}$ for the effect event ee_1.

Experiments

Dataset and Evaluation

We verify the effectiveness of the proposed model on the Chinese benchmark dataset ECE-CCKS (Tianchi 2021), which contains 7,000 sentences, 15,816 events, 7,908 cause-effect event pairs, 39 types of events and 3 types of event roles, namely Region, Industry, and Product.

Following Cui et al. (2022), we divide the dataset with the proportion of 8:1:1 and evaluate our model using Precision (P), Recall (R) and Micro-F1 (F1) of three metrics. 1) **EAE Metric**, which evaluates the model's ability to extract event arguments. 2) **CET Metric**, which evaluates the model's ability to identify cause-effect types. 3) **ECE Metric**, which synthesizes the above two metrics, where an argument in ECE is correctly extracted when its predicted cause-effect event type, span and event role simultaneously meet the gold label.

Implementation Details

We use HuggingFace's Transformers library to implement the BERT base model and MacBERT base model. The Adam algorithm (Kingma and Ba 2015) is used as an optimizer, the learning rate is initialized to 5e-5, the weight β is set to 0.1, the R-GAT layers is set to 2, the number of attentional heads Q is set to 4, the num of relational heads K is set to the number of relationships, and the dimension of the relation embedding $r_{i,j}$ is set to 30, which is randomly initialized via sampling from a uniform distribution and is learned together with the model training process. Moreover, since the positive samples in the dataset are sparse, we adopt a negative sampling rate of 0.6 for training.

Baselines

We compare our model with the following baseline methods, which can be divided into two categories depending on whether utilizing the Large Language Model (LLM).

Methods without LLM: 1) **Novel-tagging**, which follows Zheng et al. (2017), designs a unified label space, and conducts ECE via sequence-labeling. 2) **CasECE**, which follows Wei et al. (2020), first extracts the cause event and then deduces the effect event. 3) **Pair-linking**, which follows Wang et al. (2020b), first conducts event-type-level pair

	EAE(%)			CET(%)			ECE(%)		
	P	R	F1	P	R	F1	P	R	F1
Novel-tagging$_{[\circ]}$	59.40	28.47	38.49	49.79	61.70	55.11	51.52	26.75	35.22
CasECE$_{[\circ]}$	36.88	36.72	36.80	58.26	59.70	58.97	31.30	41.81	35.80
Pair-linking$_{[\circ]}$	47.08	46.49	46.79	55.78	62.95	59.14	39.24	47.69	43.05
DualCor$_{[\circ]}$	58.05	47.60	52.31	61.75	58.19	59.92	48.56	44.85	46.63
DualCor$_{[\diamond]}$	67.64	49.19	56.96	**70.68**	60.95	65.45	**58.29**	48.02	52.66
ChatGLM-6B	**68.92**	55.10	**61.24**	69.08	62.07	65.39	57.68	50.91	54.08
JEF-HM$_{[\circ]}$	60.04	53.36	56.51	63.05	64.08	63.56	50.94	51.73	51.33
JEF-HM$_{[\diamond]}$	61.49	**55.66**	58.43	66.41	**66.08**	**66.24**	53.90	**54.76**	**54.34**

Table 2: Experimental results on the ECE-CCKS dataset, where [$\circ$] and [$\diamond$] denote models that use BERT (Devlin et al. 2019) and MacBERT (Cui et al. 2020) as encoders, respectively.

linking to derive the cause-effect types, which is then used as conditional information for word-pair linking to derive event arguments. 4) **DualCor** (Cui et al. 2022), which designs a dual grid tagging scheme and explores the intra-event and inter-event arguments correlation for the task.

Methods with LLM: ChatGLM-6B (Zeng et al. 2023), is an open bilingual language model based on General Language Model (GLM) (Du et al. 2022) framework and can performs well on Chinese corpus. We implement it by finetuning on the ECE-CCKS dataset.

Overall Results

Table 2 shows the overall experimental results. The following can be observed from the table:

1) Our method JEF-HM outperforms all the methods without LLM in the F1-score, indicating that JEF-HM is effective for the ECE task. Specifically, compared with the current best method DualCor$_{[\diamond]}$, our method JEF-HM$_{[\diamond]}$ achieves 1.47%, 0.79%, and 1.68% improvements in the F1-score on three metrics, respectively. This can be attributed to that the baseline methods usually use two independent grids for cause events and effect events, and mainly focus on the relationships of “event-argument” and “argument-argument”, while we further consider the relationship of “event-event”, and the joint framework we designed can effectively capture the dependency between the two subtasks.

2) Compared with the ChatGLM-6B, our method JEF-HM$_{[\diamond]}$ achieves 0.85% and 0.26% improvements in the F1-score on CET and ECE, respectively. However, ChatGLM-6B performs better than existing methods on EAE. This can be attributed to that the ECE-CCKS dataset only considers three common roles “Region”, “Industry”, and “Product” for each event, while the ChatGLM-6B we used is a pre-trained model with rich and general knowledge, and we further finetuned it on the ECE-CCKS dataset, which allows it to perform well on traditional EAE task. However, it is worth noting that ChatGLM-6B still has certain limitations in its reasoning abilities, such as logical reasoning and relational inference, which consequently restrict its performance on CET task, while we consider multiple relationships between events and arguments, which is helpful to capture potential causal cues.

	EAE(%)	CET(%)	ECE(%)
JEF-HM$_{[\diamond]}$	58.43	66.24	54.34
-HRAG	57.72	65.24	53.10
-$\mathcal{E}_{E-E}$	56.77	63.95	52.72
-$\mathcal{E}_{E-A}$	57.06	64.25	52.61
-$\mathcal{E}_{A-A}$	57.13	64.34	52.62
-MCLE	55.17	63.80	51.38
-ES	56.53	65.18	52.96
-$\mathcal{L}_{lr}$	58.20	65.03	53.20

Table 3: Ablation study (F1/%) on the ECE-CCKS dataset.

Ablation Study

This section analyzes the contribution of each part in our model through ablation experiments, as shown in Table 3.

The effect of heterogeneous-relation-aware graph module. We use -HRAG, -$\mathcal{E}_{E-E}$, -$\mathcal{E}_{E-A}$, and -$\mathcal{E}_{A-A}$, denote the removal of the whole graph module and each type of relationship, respectively. The performance has dropped to varying degrees on three metrics, indicating that all three relationships we considered are beneficial to the model. Moreover, we note that the performance of removing each relationship on the heterogeneous graph is lower than that of removing the whole graph. This could be due to the relatively small size of our dataset, which may prevent the graph model from learning the optimal representations.

The effect of multi-channel label enhancing module. We use -MCLE and -ES denote the removal of the whole module and the enhancing strategy, respectively. After removing them, the performance on three metrics is significantly dropped, beacuse this module can help the model better learn the distributed representation of each label in multi-label framework and the enhancing strategy can leverage the preliminary results of CET and EAE as prior information to enhance the interaction between two subtasks.

The effect of auxiliary supervision. we use -$\mathcal{L}_{lr}$ denotes the removal of the auxiliary supervision for the multi-channel label representations. The auxiliary supervision can effectively constrain the distributed representation of each label, while the absence of such supervision impacts its representations, thereby affecting the overall performance.

Sentence: The recent sharp decline in international natural gas prices has led to a significant increase in the supply capacity of synthetic ammonia in the international market, thus causing a rapid decline in international urea market prices.

	Cause-Effect Type	Cause Event Arguments	Effect Event Arguments
Gold	<Price Declination, Supply Rise>	“international”$_{Region}$, “natural gas”$_{Product}$	“international”$_{Region}$, “synthetic ammonia”$_{Product}$
	<Supply Rise, Price Declination>	“international”$_{Region}$, “synthetic ammonia”$_{Product}$	“international”$_{Region}$, “urea”$_{Product}$
DualCor$_{[\diamond]}$	<Price Declination, Price Rise>		“international”$_{Region}$
	<Price Declination, Price Declination>	“international”$_{Region}$, “synthetic ammonia”$_{Product}$	“international”$_{Region}$, “urea”$_{Product}$
ChatGLM-6B	<Price Declination, Supply Rise>	“international”$_{Region}$, “natural gas”$_{Product}$	“international”$_{Region}$, “synthetic ammonia”$_{Product}$
JEF-HM$_{[\diamond]}$	<Price Declination, Supply Rise>	“international”$_{Region}$, “natural gas”$_{Product}$	“international”$_{Region}$, “synthetic ammonia”$_{Product}$
	<Supply Rise, Price Declination>	“international”$_{Region}$, “synthetic ammonia”$_{Product}$	“international”$_{Region}$, “urea”$_{Product}$

Figure 6: The outputs of different models for a given sentence.

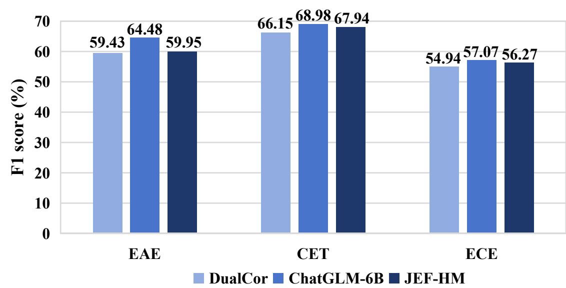


Figure 4: Experimental results on the single pair subset.

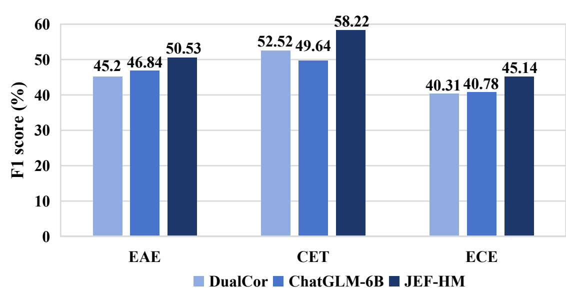


Figure 5: Experimental results on the multi pairs subset.

Single pair vs. Multi pairs

To further explore the effectiveness of our model, we divide the test set into a single pair subset and a multi pairs subset according to the number of cause-effect event pairs in the sentence, where the proportion of sentences with multiple cause-effect event pairs is 12.71%. We report the performance of three models DualCor$_{[\diamond]}$, ChatGLM-6B and JEF-HM$_{[\diamond]}$ on the two subsets, respectively, as shown in Figure 4 and Figure 5. We can observe that ChatGLM-6B performs better than the other models on the single pair subset, however, on the multi pairs subset, the performance of our model is greatly improved. This finding indicates that ChatGLM-6B has a strong ability to solve the simple samples, while its ability to solve the complex samples is relatively limited. However, our model JEF-HM$_{[\diamond]}$ can effectively improve the performance on the multi pairs subset by considering multiple relationships between events and arguments, as well as interaction information between subtasks.

Case Study

In this section, we conduct a case study to further illustrate an intuitive impression of our method. Figure 6 shows the ground truth label and the predicted results of DualCor$_{[\diamond]}$, ChatGLM-6B, and JEF-HM$_{[\diamond]}$ for the given sentence. We can observe that DualCor$_{[\diamond]}$ fails to accurately identify the cause-effect types, while ChatGLM-6B only extracts one cause-effect event pair. In contrast, our model accurately extracts all cause-effect event pairs. This case once again demonstrates the effectiveness of our method on the multi pairs subset.

Conclusion

In this paper, we propose a joint multi-label extraction framework with heterogeneous-relation-aware graph and multi-channel label enhancing strategy for the ECE task. The experimental results on the benchmark dataset ECE-CCKS indicate that our approach is effective for the ECE task, and that our model also performs well on the complex samples with multiple cause-effect event pairs. In the future, we aim to expand the dataset and mine other potential causal features for ECE task. Moreover, we only compare with the LLM ChatGLM-6B, in the next work, we will consider leveraging the advantages of the LLM (e.g., its good performance on EAE) to assist the ECE task.

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 Key Research and Development Program of China (2022QY0300-01), National Natural Science Foundation of China (62106130, 62376143, 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

Cao, P.; Zuo, X.; Chen, Y.; Liu, K.; Zhao, J.; Chen, Y.; and Peng, W. 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)*, 4862–4872.

Che, W.; Feng, Y.; Qin, L.; and Liu, T. 2021. N-LTP: An Open-source Neural Language Technology Platform for Chinese. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 42–49. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Chen, H.; Zhai, Z.; Feng, F.; Li, R.; and Wang, X. 2022a. Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2974–2985. Dublin, Ireland: Association for Computational Linguistics.

Chen, M.; Cao, Y.; Deng, K.; Li, M.; Wang, K.; Shao, J.; and Zhang, Y. 2022b. ERGO: Event Relational Graph Transformer for Document-level Event Causality Identification. In *Proceedings of the 29th International Conference on Computational Linguistics*, 2118–2128. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Cui, S.; Sheng, J.; Cong, X.; Li, Q.; Liu, T.; and Shi, J. 2022. Event Causality Extraction with Event Argument Correlations. In *Proceedings of the 29th International Conference on Computational Linguistics*, 2300–2312. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; and Hu, G. 2020. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 657–668. Online: Association for Computational Linguistics.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 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)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Dozat, T.; and Manning, C. D. 2017. Deep Biaffine Attention for Neural Dependency Parsing. In *International Conference on Learning Representations*.

Du, Z.; Qian, Y.; Liu, X.; Ding, M.; Qiu, J.; Yang, Z.; and Tang, J. 2022. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 320–335.

Gao, L.; Choubey, P. K.; and Huang, R. 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 Papers)*.

Hashimoto, C. 2019. Weakly Supervised Multilingual Causality Extraction from Wikipedia. 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)*, 2988–2999. Hong Kong, China: Association for Computational Linguistics.

Hu, Z.; and Walker, M. 2017. Inferring Narrative Causality between Event Pairs in Films. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 342–351.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *the 6th International Conference on Learning Representations*.

Liu, J.; Chen, Y.; and Zhao, J. 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*, 3608–3614.

Ning, Q.; Feng, Z.; Wu, H.; and Roth, D. 2018. Joint Reasoning for Temporal and Causal Relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2278–2288.

Riaz, M.; and Girju, R. 2014. In-depth exploitation of noun and verb semantics to identify causation in verb-noun pairs. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 161–170.

Shen, S.; Zhou, H.; Wu, T.; and Qi, G. 2022. Event Causality Identification via Derivative Prompt Joint Learning. In *Proceedings of the 29th International Conference on Computational Linguistics*, 2288–2299. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Shi, J.; Cao, S.; Hou, L.; Li, J.; and Zhang, H. 2021. TransferNet: An Effective and Transparent Framework for Multi-hop Question Answering over Relation Graph. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 4149–4158. Online and Punta

Cana, Dominican Republic: Association for Computational Linguistics.

Tianchi. 2021. CCKS2021 the Dataset for Financial Event and Causal Relation Extraction.

Wang, K.; Shen, W.; Yang, Y.; Quan, X.; and Wang, R. 2020a. Relational Graph Attention Network for Aspect-based Sentiment Analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3229–3238. Online: Association for Computational Linguistics.

Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; and Sun, L. 2020b. TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking. In *Proceedings of the 28th International Conference on Computational Linguistics*, 1572–1582. Barcelona, Spain (Online): International Committee on Computational Linguistics.

Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; and Chang, Y. 2020. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1476–1488. Online: Association for Computational Linguistics.

Wu, S.; Zhao, R.; Zheng, Y.; Pei, J.; and Liu, B. 2023. Identify Event Causality with Knowledge and Analogy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 13745–13753.

Zeng, A.; Liu, X.; Du, Z.; Wang, Z.; Lai, H.; Ding, M.; Yang, Z.; Xu, Y.; Zheng, W.; Xia, X.; Tam, W. L.; Ma, Z.; Xue, Y.; Zhai, J.; Chen, W.; Liu, Z.; Zhang, P.; Dong, Y.; and Tang, J. 2023. GLM-130B: An Open Bilingual Pre-trained Model. In *The Eleventh International Conference on Learning Representations (ICLR)*.

Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; and Xu, B. 2017. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1227–1236. Vancouver, Canada: Association for Computational Linguistics.

Zuo, X.; Cao, P.; Chen, Y.; Liu, K.; Zhao, J.; Peng, W.; and Chen, Y. 2021. 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)*, 3558–3571.

Zuo, X.; Chen, Y.; Liu, K.; and Zhao, J. 2020. KnowDis: Knowledge Enhanced Data Augmentation for Event Causality Detection via Distant Supervision. In *Proceedings of the 28th International Conference on Computational Linguistics*, 1544–1550.