Tech Science Press

Exploiting Rich Event Representation to Improve Event Causality Recognition

Gaigai Jin[1], Junsheng Zhou[1,*], Weiguang Qu[1], Yunfei Long[2] and Yanhui Gu[1]

[1]School of Computer and Electronic Information, Nanjing Normal University, Nanjing, 210046, China
[2]School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK
*Corresponding Author: Junsheng Zhou. Email: zhoujs@njnu.edu.cn

Abstract: Event causality identification is an essential task for information extraction that has attracted growing attention. Early researchers were accustomed to combining the convolutional neural network or recurrent neural network models with external causal knowledge, but these methods ignore the importance of rich semantic representation of the event. The event is more structured, so it has more abundant semantic representation. We argue that the elements of the event, the interaction of the two events, and the context between the two events can enrich the event's semantic representation and help identify event causality. Therefore, the effective semantic representation of events in event causality recognition deserves further study. To verify the effectiveness of rich event semantic representation for event causality identification, we proposed a model exploiting rich event representation to improve event causality recognition. Our model is based on multi-column convolutional neural networks, which integrate rich event representation, including event tensor representation, event interaction representation, and context-aware event representation. We designed various experimental models and conducted experiments on the Chinese emergency *corpus*, the most comprehensive annotation of events and event elements, enabling us to study the semantic representation of events from all aspects. The extensive experiments showed that the rich semantic representation of events achieved significant performance improvement over the baseline model on event causality recognition, indicating that the semantic representation of events plays an important role in event causality recognition.

Keywords: Event tensor representation; event interactive representation; context-aware event representation; event causality identification; multi-column convolutional neural networks

1 Introduction

Understanding events is an important component of natural language understanding. An essential step in this process is identifying causality between events. Event causality identification is an important natural language processing task that can benefit question answering [1], reading comprehension, event prediction, narrative generation, and financial analysis [2].

Event causality identification was first inspired by linguistics. Early researchers identified causal connectives in event text. Additionally, some researchers presented the novel joint framework of Temporal and Causal Reasoning (TCR), which regards event causality identification as an integer linear programming with constraints and linguistic rules. The method combining the neural network and external causal knowledge has shown excellent performance.

However, there still exist some deficiencies in event causality identification. First, the present models only base performance on simple phrase structures while ignoring the structured event texts with multiple attributes. Simple word embedding cannot clearly distinguish event semantics, although the event representation calculated by a tensor has shown to be effective in event discrimination and event prediction/generation. The Role Factorized Tensor (RFT) model [3] provides multiplicative interactions between predicates and attributes. It captures different scenarios (or contexts) that use the same predicate and can obtain the effective semantics of an event.

Second, existing event causality recognition models only encode single event text while ignoring the interaction between events. People can get the semantic association between two events by reading an article repeatedly. Interactive attention (IA) [4] can imitate people's reading habits and obtain the interactive semantic representation of events by calculating the interactive attention between two event texts. The interactive event representation can get the dependency semantics between the two events and help identify the relationship between events.

However, two events with causality are not always adjacent in the text. Most of the existing models only focus on identifying causal relations between events in sentences. These methods ignore the context between the events across sentences, resulting in logic loss and semantic loopholes. To obtain an effective representation of the event, we should consider the context between event pairs. Gate Attention (GA) [5] was first applied to reading comprehension tasks. Specifically, the GA allows the query to directly interact with each dimension of the token embedding at the semantic-level. This ability can obtain finer, more accurate semantics within the text. Context-aware representation of events calculated by GA can make up for the semantic deficiency of event representation.

Multi-column convolutional neural networks (MCNNs) [6] are a variant of convolutional neural networks [7] with several independent columns. Each column has its own convolutional and pooling layers, and the outputs of all of the columns are combined in the last layer to get a final prediction. In order to make up for the shortcomings of existing research methods, we proposed an event causality identification model based on the multi-column neural networks model, which integrates three different event representations:

(1) Event tensor representation: We first integrate the Role Factorized Tensor model to get different scenarios or context event semantics in which we use the same predicate;

(2) Interactive event representation: We get the interactive semantics between two events through Interactive Attention, which can obtain the dependency semantics between the two events and can identify the relationship between the two events;

(3) Event representation with context awareness can complement obtain event semantics of long-distance event pairs and reduce semantic loopholes. Gated attention can get fine-grained semantic content and aid in event representation.

The rest of the paper is organized as follows: Section 2 introduces the relevant work, and then we derive the basic model of the experiment by dong improving on an off-the-shelf model. In the core section (Section 4), we describe three types of event representations: event tensor representation, event interaction representation, and discourse-aware event representation.

2 Related Work

In this section, we briefly review two research areas related to our work: event causality identification and event representation.

2.1 Event Causality Identification

Previous researchers exploited various clues for event causality recognition, including discourse connectives and word sharing between cause and effect. The works of Rahimtoroghi et al. [8] and Do et al. [9] focused on the causal association degree (Causal Effect Association or Causal Potential) by calculating the multiple-point mutual information of the event trigger and other elements of the event or used the principle of event co-occurrence to generate causal pairs.

Some researchers formulated the problem as a classification task and determined the label of each pair locally. For example, Zhao et al. [10] used Bayesian networks to identify the causality between Aerospace events according to the probability of occurrence of causal events.

Other researchers posited that the event temporal relation and causality are often influenced by each other and proposed a joint learning method to identify the event's temporal relation and event causality simultaneously. They [11–13] regarded the event causality identification as integer linear programming with constraints and linguistic rules. The joint recognition model proposed by Zhang et al. [14] takes the event temporal as the main task and the event causality as the auxiliary task, and uses two neural network encoding layers to realize it.

With the extensive application of deep learning, combining neural networks and external causal knowledge has achieved excellent performance. Li et al. [15] used prior knowledge of vocabulary knowledge. The keywords representing the causal relationship from the sentence were used to identify the text's causal relationship. Kruengkrai et al. [16] proposed multi-column convolution to combine three kinds of external knowledge (the answer of question-answer system, one or two consecutive sentences with causal trigger and causality template), and they divided the original text into five segments to identify causal event pairs such as "smoking" → "death or lung cancer". Furthermore, Kadowaki et al. [17] proposed a BERT-based method that exploits each annotator's independent judgments to recognize event causality.

2.2 Event Representation

Experts introduced the concept of event representation because event text differs from raw text. It is necessary to represent event semantic for many natural language comprehension tasks. Most previous studies used discrete event representation. They defined an event as a 6-tuple $e = (P, O_1, O_2, O_3, O_4, t)$ [18] where P is the action or state that marks the occurrence of the event, t is the time of the event, $O_1, O_2, O_3,$ and O_4 represent the implementer of one or more events, the object that one or more events act on, one or more tools that make the event occur, and one or more places, or $e = (A, O, T, P, S, L)$ [19], which represent action, object, time, place, status, language expression, respectively.

However, discrete event representation is sparse. Therefore, experts began to use a deep neural network to learn dense vector representation of events to alleviate the sparsity. It is necessary to design appropriate training objectives for event representation to make the embedded dense event representation vector retain rich semantic information. The methods of learning event representation include using the tensor calculation method to get event representation based on the structural information of the event itself [20], obtaining event representation based on the interaction semantics between events [21], and integrating external causal knowledge into event representation [22–24].

Although there has been ample research on event causality identification and event representation, there is no precedent for integrating rich event representation into the event causality identification model.

Therefore, we integrated multiple representations of the event into the multi-column neural networks model to improve event causality recognition in our work.

3 Baseline Model

At present, the combination of neural networks and external knowledge is the most widely used method in event causality identification. Better results are based on the multi-column convolutional neural networks and the Siamese network. The multi-column convolutional neural networks can increase the use of external knowledge while acquiring rich event semantics. However, the convolutional neural network can only obtain n-gram semantic information, while the recurrent neural network can get long-distance semantic information. Thus, we spliced the convolution layer and Bi-GRU layer to obtain multi-level event semantic information.

We implemented the baseline model based on the multi-column convolutional neural networks (Fig. 1). Multi-column convolutional Bi-GRUs contain three neural networks with the same parameters. Compared with a single neural network, the multi-column convolutional neural networks can extract diverse features. Moreover, we add Bi-GRU units to obtain long-distance event semantic information.
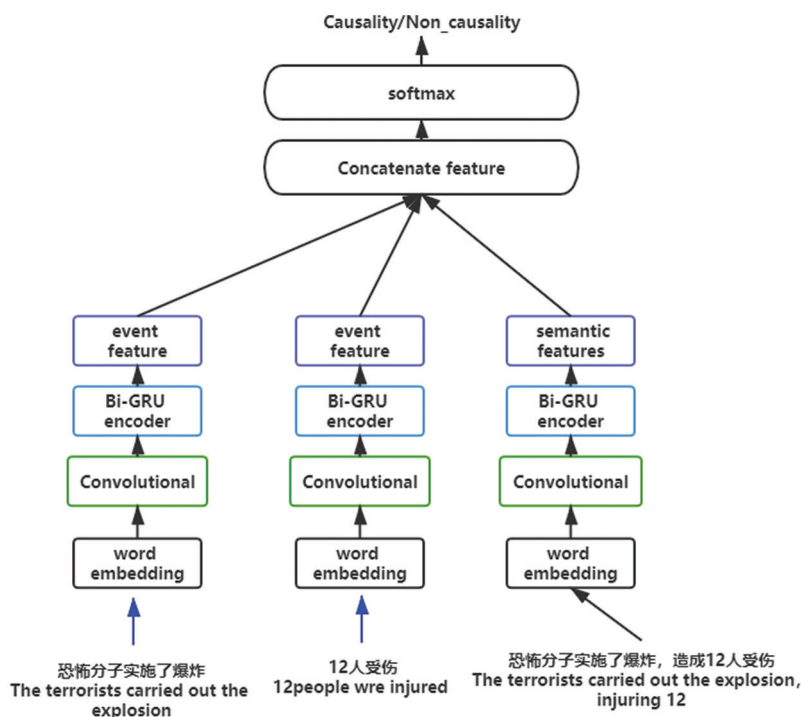

Figure 1: Baseline Model

We extract events from news according to event triggers. Then, we get two events and an event pair. These fragments are input as contexts into the three columns. After we obtain the layer and Bi-GRU encoder, we can obtain more semantic information of the event. The single event semantic features and event pair semantic are concatenated and sent to the classifier's final classification layer.

4 Exploiting Rich Event Representation

Most existing methods are based on external causal knowledge and neural network. The common problem is that they only focus on the plane semantic representation of events while ignoring the rich semantic representation of events. To solve this problem, we propose an event causality recognition model exploiting rich event representation.

Our model is based on the multi-column neural network and contains three types of event representation: event tensor representation, event interactive representation, and context-aware event representation. The rest of this section describes these three event representations in detail.

4.1 Incorporating Event Tensor Representation

Event tensor representation is based on the Role Factored Tensor Model (RFT). This model captures multiplicative interactions across all event arguments, which allows the predicate tensor to model complex relations between the predicate and its arguments. The method yields representations that are effective at event causality recognition because it captures sensitive semantic interactions between an event and its entities and changes in the event's components should drastically change the value of the event embedding. A model employing the tensor representation of the event is shown in Fig. 2.
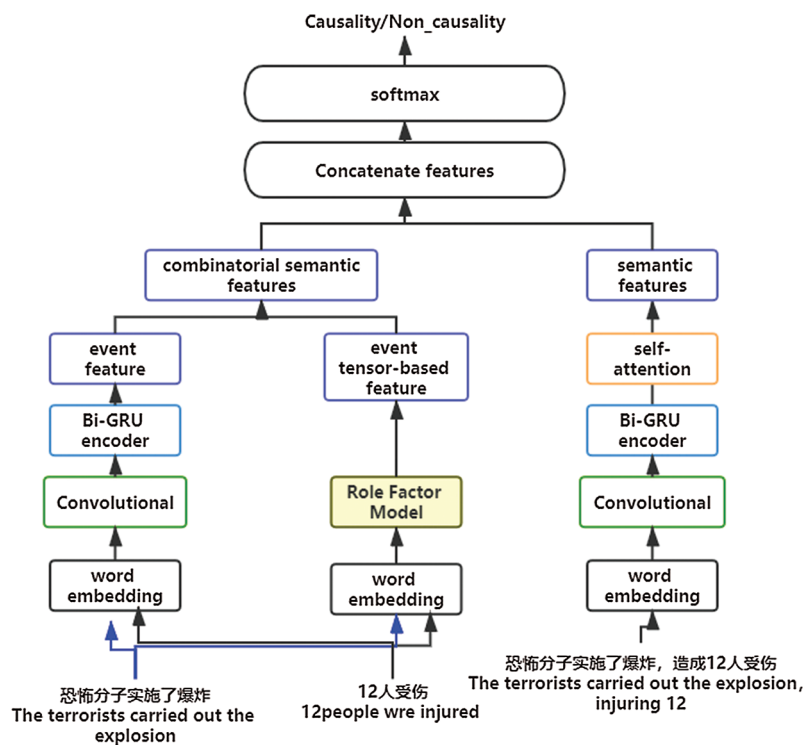
Figure 2: Event tensor representation model

Compared with ordinary text, the event is more structured and includes an event trigger and type, participants, time, location, and other event elements. Therefore, event representation must consider the impact of each element on the final semantic representation of the event. For example:

例 1: 恐怖分子实施了爆炸, 造成12人受伤, 3人死亡.

Example 1: The terrorists carried out the explosion, injuring 12 and killing three.

例 *2:* 救援小组实施抢险救援，受伤人员得到及时救助.

Example 2: The rescue team carried out an emergency rescue and got out the injured in time)

In the above two examples, the event semantics of the same trigger 实施 (*carried out*) are completely different when matched with different arguments. The results in *Example 1* and the result in *Example 2* are different because of this difference.

The significance of tensor computing in knowledge graph tasks demonstrates its ability to model data interactivity and obtain the semantics between words or phrases. The RFT model is a method for calculating the event representation of a Bi-Linear tensor based on event role decomposition. We compute the event trigger and event arguments to obtain the event tensor representation.

The original neural network is used to extract the event text's flat features, and we can obtain two kinds of event representations. These two kinds of event representations are concatenated to help identify causality between events. For an event pair, the convolution layer captures n-gram semantic information, the Bi-GRU layer obtains long-distance context semantic representation, and the self-attention layer assigns different weights to each element according to its contribution, which can determine the attention to the specific features of the event.

For a single event, a pre-trained word vector GloVe is used to represent the semantic of a single event, and the input of the other columns are the event trigger, arguments, and other elements. These can obtain the event tensor representation through the RFT model.

Event tensor representation uses the RFT model to capture interactions between the event trigger and its arguments separately to combine them into the final embedding. The argument-specific interactions are captured using two compositions of the trigger: one with the subject and one with the object:

$$v_s = T(s, d), v_o = T(o, d), \tag{1}$$

where S is the subject, O is the object, and D is the event trigger "denote".

$$T(A, B) = f(A^T W_1 B + W_2[A, B] + b), com_{rep} = W_s v_s + W_o v_o. \tag{2}$$

The input event text vector includes word embedding and part-of-speech (POS) tagging information of event text, $x_i = x_w \oplus x_p$, where x_w is word embedding, and x_p is the POS of the word.

4.2 Integrating Interactive Event Representation

We show the model that integrates interactive event representation in Fig. 3. Interactive Attention (IA) can get the interactive semantic between two events by simulating people's reading habits back and forth. It can provide effective semantics for identifying the relationship between two events.

After integrating the event tensor representation calculated by RFT to the multi-column neural network model, the experimental results improved, confirming the effectiveness of event semantic representation for event causality identification. However, the event tensor representation only uses the event's elements, and the interaction between events also needs to be considered. In other words, the representation between events should include the interactive semantics of two events, which represents the interaction between two events. The interactive event representation contains more semantic information than the event representation, which only depends on the word embedding.

Considering that the events with causality in the news are not independent, the two events with causality have strong semantic relevance. When identifying the causality between two events, good event representation should keep the element information of the event and include the interactive semantics between event pairs. Attention mechanism has achieved success in many NLP tasks and IA can imitate the human-like bidirectional reading strategy and enhance event representations.
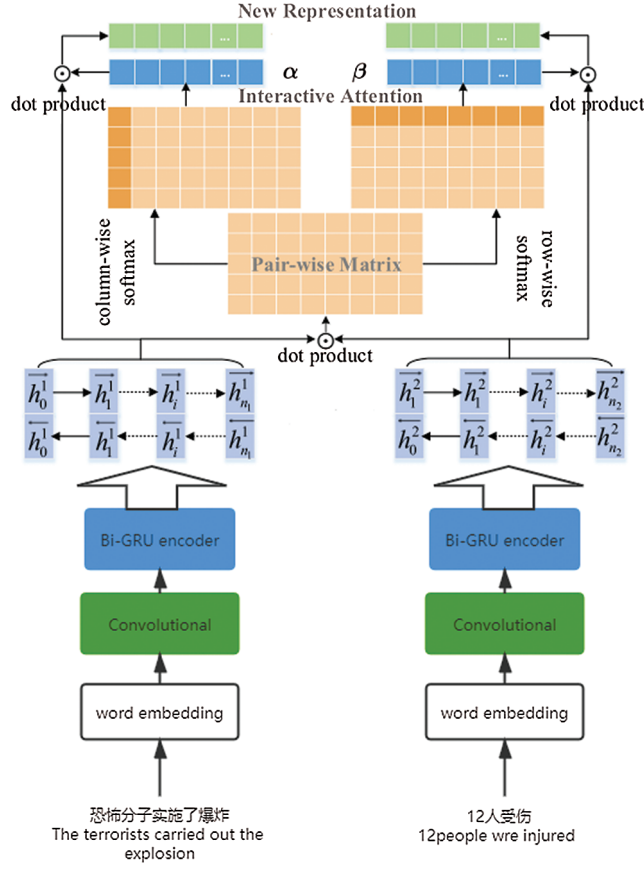
Figure 3: Interactive event representation framework

First, we can get a pair-wise matrix through a calculation expressing the semantic relationship between the word pairs of two events $S(i,j) = h_i^1 T h_j^2$. $S(i,j)$ is the correlation score between the i-th word of *Event1* and the j-th word of *Event 2*. The matrix is a matrix of $n_1 * n_2$ derived by the dot product of their hidden layer representation, where n_1, n_2 represents the text length of two events. Second, when judging the relationship between the two events, the attention is asymmetrical when reading in order and reverse order. For each word in *Event2*, normalization applies to each column on matrix S to obtain its probability distribution on *Event1* $\alpha_t \in \Re^{n_1}$. Similarly, normalization is applied to each row on the matrix S to obtain the attention distribution of each word in *Event1* on *Event2* $\beta_t \in \Re^{n_2}$.

$$\alpha_t = [\alpha_t^1, \alpha_t^2, \ldots, \alpha_t^{n_1}], \quad \alpha_t^i \text{ expresses the attention of the } i-\text{th word of Event1 on position t.} \tag{3}$$

$$\beta_t = [\beta_t^1, \beta_t^2, \ldots, \beta_t^{n_1}], \quad \beta_t^i \text{ expresses the attention of the } j-\text{th word of Event2 on position t.} \tag{4}$$

In order to express the interactive semantic connection between two events, we use all α_t, β_t to obtain the final averaged attention α, β of *Event1* and *Event2*:

$$\alpha = \frac{1}{n_2} \sum_{t=1}^{n_2} \alpha_i, \beta = \frac{1}{n_1} \sum_{t=1}^{n_1} \beta_i \tag{5}$$

The interactive event representations integrating event context and interactive attention are shown in Eq. (6), which reflects the human-like bidirectional reading strategy to some extent:

$$h_{event1} = h^1 * \alpha, h_{event2} = h^2 * \beta. \tag{6}$$

4.3 Exploiting Context-Aware Event Representation

Context-aware event representation can overcome the semantic default and logic loopholes in distant event pairs. Compared with ordinary attention, Gated Attention (GA) can obtain finer and more accurate semantics of the text, and the context-aware representation of an event calculated by GA can complete the semantics of distant event pairs.

Existing research methods focus on analyzing two separate events; they lack context between events. However, Chinese is full of flexible sentence structures and semantic cohesion, such as ellipses, references, substitutions, and conjunctions. Simultaneously, the two events' position in the text is not always adjacent, and there is often semantic deviation when analyzing two events separately. Therefore, the causality recognition of distant event pairs must consider text content's influence on event semantic representation.

GA allows an event to directly interact with each dimension of the token embedding in context at the semantic level.

First, we use two events as endpoints to get the text content, and we divide it into three parts: start event, text content, and end event. Then, the gate attention calculation is performed on the start event and the end event with the text content to obtain the context-aware event representation.

For each word in context D, GA uses soft attention to construct the representation vector in Event representation, and then uses this vector to multiply the representation vector of the context,

$$\alpha_i = soft \max(E^T d_i), \tag{7}$$

$$e_i = Q\alpha_i, x_i = d_i \cdot e_i, \tag{8}$$

where context representation is X, the event representation of context awareness is $E'_{text} = C(X, E)$, and $C(A, B) = conc(A, B) || add(A, B) || dot\_product(A, B)$

Finally, we combined the three semantic event representations, and combinatorial features $event\_rep = C(com_{rep}, Inter_{rep}, context - aware_{rep})$ are predicted by a final classifier. We show our model in Fig. 4. The objective function of the model is defined as $loss\_cross\_entrophy(Y, \hat{Y})$.
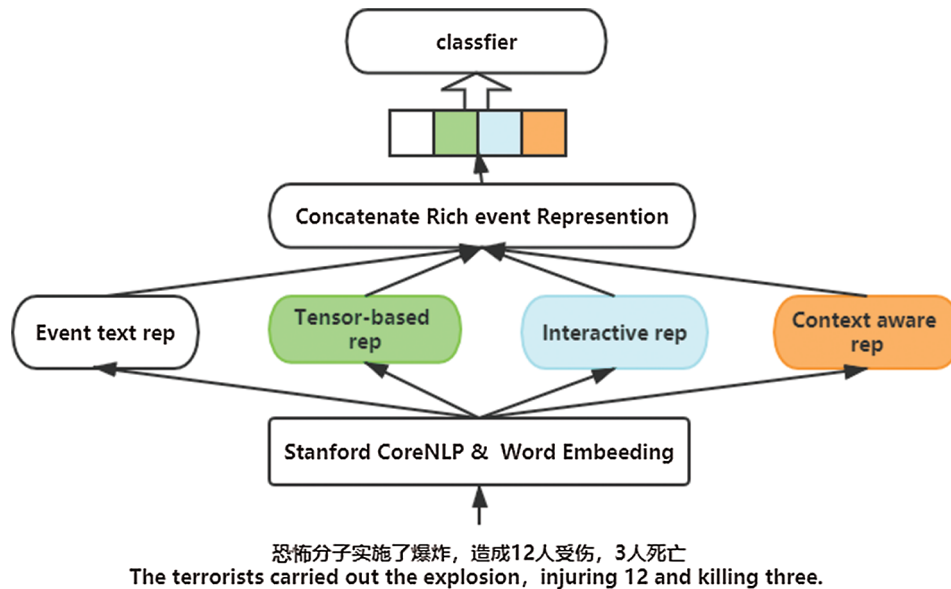

Figure 4: Simplified overall architecture

5 Experiments

5.1 Data Set and Evaluation

Data set. In our experiments, we selected CEC2.0 [25] (Chinese emergency *corpus*) as the experimental data set, which was constructed by the Semantic Intelligence Laboratory of Shanghai University. The *corpus* included 332 news reports on five types of emergencies (earthquake, fire, traffic accident, terrorist attack, and food poisoning) collected from the Internet. The raw *corpus* was then processed by text preprocessing, text analysis, event tagging, and consistency checking. Finally, the annotation results were the *corpus*.

CEC2.0 adopts XML as the annotation format, which contains six data structures (tags): event, reminder, time, location, participant, and object. Event is used to describe the event; trigger (denoter), time, location, and participant, and objects are used to describe the event's indicators and elements. Compared with ACE and TimeBank, CEC2.0 is smaller in scale but has the most comprehensive annotation of events and event elements. The CEC2.0 data set labeled each attribute of events clearly. Moreover, the CEC2.0 *corpus* contains a variety of event relationships. In the experiments, we only identified whether there was a causal relationship between events.

We segmented the data set according to the ratio of 7:1:2 (as shown in Tab. 1) and obtained 1914 training sets, 273 verification sets, and 548 pieces of test data.

Table 1: Experimental data set distribute

Data set	Causal	Non-Causal
Train	655	1264
Val	100	173
Test	207	345
Total	962	1782

Evaluation Metrics. We employed Accuracy (Acc.), Precision (P), Recall (R), and F1 (F) as our main evaluation metrics.

5.2 Experimental Parameter Setting

We used the Stanford CoreNLP tool to segment event texts and GloVe to train word vectors on the training data and event representation. The dimension of the word vector was 100, and the hidden size was 128. We chose Adam as the optimizer, and the learning rate was 1e-3 (as shown in Tabs. 2 and 3).

Table 2: Parameters of baselines

Parameters	Value
Embedding dim	100
Hidden size	64
Kernel size	[3,4,5]
Loss function	Cross-Entropy Loss

Table 3: Parameters of all other models

Parameters	Value
Embedding dim	100
Hidden size	64
Optimizer	Adam
Learning rate	1e-3
Loss function	Cross-Entropy Loss

5.3 Baseline Model

We chose several commonly used neural network models to identify a causal relationship between events, including MCNN, Siamese Bi-LSTM, and CNN. We also attempted to add attention to the deep neural network. The multi-column convolution Bi-GRU model performed better (as shown in Tab. 4).

Finally, we chose MCGRU (Multi-column Convolution and Bi-GRU) as the baseline model to identify event causality. The experimental results showed that the composition of convolution and Bi-GRU could obtain more efficient event semantics, which could help to identify event causality.

Table 4: Baseline models' results

Model	Accuracy	Precision	Recall	F1-score
CNN	85.95	80.43	74.75	77.49
MCNN	88.24	80.58	83.64	82.08
Siamese Bi-LSTM (α = 0.3)	–	82.51	81.62	82.07
MBi-GRU	87.91	78.18	86.87	82.30
MAtt-CGRU	88.56	82.00	82.83	82.41
MCGRU	**88.56**	**80.19**	**85.84**	**82.93**

5.4 Results and Discussions

Tab. 5 shows the recall, precision, F1 score, and accuracy of the test data set under each model test. MCGRU+EC+inter_att+gate_att indicates the overall model, which demonstrated the highest accuracy and the highest F1 value. MCGRU+EC+ inter_att is a model integrating interactive semantics and event tensor representation, and it had the highest accuracy, MCGRU+EC only contains the tensor of events, and it had a higher recall rate.

Table 5: Performance results of all models on CEC2.0

Model	Accuracy	Precision	Recall	F1-score
MCGRU	88.56	80.19	85.84	82.93
Siamese Bi-LSTM (α = 0.3)	–	82.51	81.62	82.07
MCGRU+EC	88.89	80.27	86.37	83.21
Siamese Bi-LSTM (α = 0.2)	–	83.01	84.65	83.82
MCGRU+EC+inter_att	89.23	**90.22**	80.19	84.91
MCGRU+EC+inter_att+gate_att	**89.60**	89.89	81.64	**85.57**

Compared with the Siamese Bi-LSTM [19], our overall model outperformed the existing model in three of the four indicators, but it had certain deficiencies in the recall rate. Therefore, we should focus on improving the recall rate of the model and increasing the accuracy rate in future work.
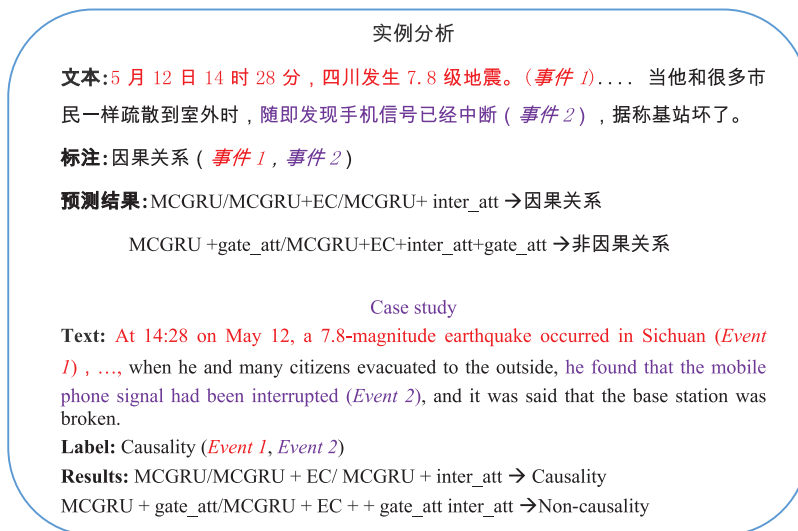
5.5 Ablation Study

In order to explore the effect of the different event representations, we conducted several ablation experiments. Here, +X adds a different event representation module (as shown in Tab. 6).

Table 6: Ablation experiment results

Model		Accuracy	Precision	Recall	F1-score
MCGRU	**No addition**	88.56	80.19	85.84	82.93
	+EC	88.89	80.27	86.37	83.21
	+inter_att	88.32	84.57	83.09	83.82
	+gate_att	88.50	83.87	82.74	83.30
MCGRU+EC+inter_att+gate_att		**89.60**	89.89	81.64	**85.57**

The ablation experiments showed that the rich event representations significantly improved the accuracy rate and precision rate. Nevertheless, the model's recall rate slightly dropped in exploiting rich event representation because the model considers rich and fine event semantic representation, resulting in overfitting.

In a future study, we should measure the influence of each event representation on the model and choose different weights for each event representation to analyze and study the effect (as shown in Fig. 5).

实例分析

文本:5 月 12 日 14 时 28 分，四川发生 7.8 级地震。(*事件 1*).... 当他和很多市民一样疏散到室外时，随即发现手机信号已经中断（*事件 2*），据称基站坏了。

标注:因果关系（*事件 1，事件 2*）

预测结果:MCGRU/MCGRU+EC/MCGRU+ inter_att →因果关系

　　　　MCGRU +gate_att/MCGRU+EC+inter_att+gate_att →非因果关系

Case study

Text: At 14:28 on May 12, a 7.8-magnitude earthquake occurred in Sichuan (*Event 1*) , …, when he and many citizens evacuated to the outside, he found that the mobile phone signal had been interrupted (*Event 2*), and it was said that the base station was broken.

Label: Causality (*Event 1*, *Event 2*)

Results: MCGRU/MCGRU + EC/ MCGRU + inter_att → Causality
MCGRU + gate_att/MCGRU + EC + + gate_att inter_att →Non-causality

Figure 5: Case study

6 Conclusions

Event causality identification is an important task for information extraction that has attracted growing attention recently. We proposed an event causality recognition model based on the multi-column neural

network that integrates event text representation, event tensor representation, event interactive representation, and context-aware event representation. The experimental results demonstrated a significant improvement in CEC2.0, which confirms that the rich representation of events plays an important role in event causality identification.

Conflicts of Interest: We declare that we have no conflicts of interest to report regarding the present study.

References

[1]   Z. Bo, W. Haowen, L. Jiang, Y. Shuhan and L. Meizi, "A novel bidirectional LSTM and attention mechanism based neural network for answer selection in community question answering," *Computers, Materials & Continua*, vol. 62, no. 3, pp. 1273–1288, 2020.

[2]   L. Shi, P. Lu and J. Yan, "Causality learning from time series data for the industrial finance analysis via the multi-dimensional point process," *Intelligent Automation & Soft Computing*, vol. 26, no. 5, pp. 873–885, 2020.

[3]   N. Weber, N. Balasubramanian and N. Chambers, "Event representations with tensor-based compositions," in *Proc. AAAI*, New Orleans, Louisiana, USA, pp. 4946–4953, 2018.

[4]   F. Guo, R. He, D. Jin, J. Dang, L. Wang *et al.,* "Implicit discourse relation recognition using neural tensor network with interactive attention and sparse learning," in *Proc. COLING*, Santa Fe, New Mexico, USA, pp. 547–558, 2018.

[5]   B. Dhingra, H. Liu, Z. Yang, W. W. Cohen and R. Salakhutdinov, "Gated-attention readers for text comprehension," in *Proc. ACL*, Vancouver, Canada, pp. 1832–1846, 2017.

[6]   D. C. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE*. Providence, RI, USA, pp. 3642–3649, 2012.

[7]   Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. IEEE*, Santa Clara, California, USA, pp. 2278–2324, 1998.

[8]   E. Rahimtoroghi, E. Hernandez and M. A. Walker, "Learning fine-grained knowledge about contingent relations between everyday events," in *Proc. SIGDIAL*, Los Angeles, USA, pp. 350–359, 2016.

[9]   Q. X. Do, Y. S. Chan and D. Roth, "Minimally supervised event causality identification," in *Proc. EMNLP*, Edinburgh, UK, pp. 294–303, 2011.

[10] S. Zhao, T. Liu and S. Zhao, "Event causality extraction based on connectives analysis," *Neuro Computing*, vol. 173, no. 3, pp. 1943–1950, 2016.

[11] Y. Huang, P. Li and Q. Zhu, "Joint model of events' causal and temporal relations identification," *Computer Science*, vol. 45, no. 6, pp. 204–207, 2018.

[12] P. Li, Y. Huang and Q. Zhu, "Global optimization to recognize causal relations between events," *Journal of Tsinghua University (Natural Science Edition)*, vol. 10, no. 4, pp. 37–42, 2017.

[13] Q. Ning, Z. Feng, H. Wu and D. Roth, "Joint reasoning for temporal and causal relations," in *Proc. ACL*, Melbourne, Australia, pp. 2278–2288, 2018.

[14] Y. Zhang, P. Li and Q. Zhu, "Joint learning for identifying temporal and causal relations between events," *Computer Engineering*, vol. 46, no. 7, pp. 65–71, 2020.

[15] P. Li and K. Mao, "Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts," *Expert Systems with Applications*, vol. 115, no. 32, pp. 512–523, 2019.

[16] C. Kruengkrai, K. Torisawa, C. Hashimoto, J. Kloetzer, J. H. Oh *et al.,* "Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks," in *Proc. AAAI*, San Francisco, California, USA, pp. 3466–3473, 2017.

[17] K. Kadowaki, R. Iida, K. Torisawa, J. H. Oh and J. Kloetzer, "Event causality recognition exploiting multiple annotators' judgments and background knowledge," in *Proc. EMNLP/IJCNLP*, Hong Kong, China, pp. 5815–5822, 2019.

[18] K. Radinsky, S. Davidovich and S. Markovitch, "Learning causality for news events prediction," in *Proc. WWW*, Lyon, France, pp. 909–918, 2012.

[19] Z. Yang, W. Liu and Z. Liu, "Event causality identification by modeling events and relation embedding," in *Proc. ICONIP*, Siem Reap, Cambodia, pp. 59–68, 2018.

[20] X. Ding, K. Liao and T. Liu, "Event representation learning enhanced with external commonsense knowledge," in *Proc. EMNLP-IJCNLP*, Hong Kong, China, pp. 4894–4903, 2019.

[21] S. Zhao, Q. Wang and S. Massung, "Constructing and embedding abstract event causality networks from text snippets," in *Proc. ACM*, Cambridge, United Kingdom, pp. 335–344, 2017.

[22] X. Ding, K. Liao and T. Liu, "Event representation learning enhanced with external commonsense knowledge," in *Proc. EMNLP*, Hong Kong, China, pp. 4894–4903, 2019.

[23] X. Ding, Y. Zhang and T. Liu, "Knowledge-driven event embedding for stock prediction," in *Proc. COLING*, Osaka, Japan, pp. 2133–2142, 2016.

[24] M. Kim, J. Kim and M. Shin, "Word embedding based knowledge representation with extracting relationship between scientific terminologies," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 141–147, 2020.

[25] J. F. Fu, "Research on event oriented knowledge processing," Ph.D. dissertation. Shanghai University, Shanghai, 2010.